

Gestión de datos genómicos con finalidad clínica y de investigación

gen&data

Grupo de Trabajo
en Gestión de Datos Genómicos

Gestión de datos genómicos con finalidad clínica y de investigación

gen&data
Grupo de Trabajo
en Gestión de Datos Genómicos

Edición electrónica navegable Optimizada para Adobe Acrobat

Si su lector de PDFs no le permite realizar esta navegación puede descargarse la aplicación gratuita Adobe Acrobat Reader haciendo [click aquí](#).

Esta publicación le permite navegar entre apartados a través de los siguientes botones:



Gestión de datos genómicos con finalidad clínica y de investigación

gen&data
Grupo de Trabajo
en Gestión de Datos Genómicos

ISBN: 978-84-944589-1-0
D.L.: B28474-2015

© del contenido: Instituto Roche.
C/ Eucalipto, 33. 28016 Madrid
www.institutoroche.es

© 2015 de esta edición: Idemm Farma S.L.
Rda. Sant Pere, 19-21, 4.º 5.ª. 08010 Barcelona
Reservados todos los derechos.

Ninguna parte de esta publicación puede ser reproducida, almacenada o transmitida en cualquier forma ni por cualquier procedimiento electrónico, mecánico, de fotocopia de registro o de otro tipo, sin el permiso previo de los autores y editores.



GRUPO DE TRABAJO EN GESTIÓN DE DATOS GENÓMICOS

Coordinadores

Pablo Lapunzina Badía

Instituto de Genética Médica y Molecular-INGEMM. Hospital Universitario La Paz. IdiPAZ. Madrid.

Pilar Nicolás Jiménez

Cátedra Interuniversitaria de Derecho y Genoma Humano. Diputación Foral de Bizkaia. Universidad de Deusto y Universidad del País Vasco. Bilbao.

Panel de Expertos

Joaquín Arenas Barbero

Instituto de Investigación del Hospital i+12. Madrid.

Javier Arias Díaz

Universidad Complutense-Hospital Clínico San Carlos. Comité de Bioética y Bienestar Animal del ISCIII. Madrid.

Ignacio Blanco Guillermo

Hospital Universitario Germans Trias i Pujol. Badalona. Barcelona.

Miguel del Campo Casanelles

Servicio de Genética. Hospital Vall d'Hebron e Instituto Dexeus. Barcelona. Department of Pediatrics, University of California, San Diego (UCSD).

Francisco Javier Fernández Martínez

Servicio de Genética. Hospital 12 de Octubre. Madrid.

Melchor Hoyos García

Hospital La Fe. Valencia.

Belén Jiménez Ojeda

Fundación Pública Andaluza Progreso y Salud. Sevilla.

Pastora Martínez Samper

Institut d'Investigacions Biomèdiques August Pi i Sunyer (IDIBAPS). Barcelona.

Jesús Rubí Navarrete

Agencia Española de Protección de Datos. Madrid.

Jose Antonio Seoane Rodríguez

Facultad de Derecho. Universidad de A Coruña.

Xosé Antón Suárez Puente

Instituto Universitario de Oncología. Universidad de Oviedo.

Nuria Terribas Sala

Fundación Víctor Grífols i Lucas. Barcelona.

Alfonso Valencia Herrera

Centro Nacional de Investigaciones Oncológicas (CNIO). Madrid.

Revisores externos

Carmen Ayuso García

Servicio de Genética. Hospital Universitario Fundación Jiménez Díaz (IIS-FJD). Madrid.

Elías Campo Güerri

Hospital Clínic. Universitat de Barcelona. Fundació Clínic per a la Recerca Biomèdica. Barcelona.



Presentación

Desde el inicio de las actividades del Instituto Roche, en el año 2004, hemos asistido al progresivo desarrollo de la Medicina Personalizada, que ya constituye una realidad en nuestro entorno, lo que ha sido posible principalmente gracias a un mayor conocimiento del genoma humano y al extraordinario avance en las tecnologías ómicas. No obstante, aún queda mucho camino por recorrer para que esté totalmente implantada en la práctica asistencial y para que sus beneficios alcancen plenamente a los pacientes.

En este sentido, la incorporación de la *Next Generation Sequencing* (NGS) o secuenciación masiva ha marcado un hito en esta notable evolución científica al hacer posible, de forma rápida y económica, la determinación del genoma completo de una persona; este hecho ha dado lugar a que, en los últimos años, la NGS se imponga paulatinamente como método estándar para la mayor parte de los procesos de secuenciación, tanto en investigación como en clínica.

La utilización de esta tecnología NGS permite obtener una enorme cantidad de datos que es preciso gestionar de forma adecuada y rigurosa, tomando en consideración múltiples aspectos que intervienen en su obtención, utilización y posterior conservación, desde los más técnicos hasta los legales y éticos.

El grupo multidisciplinar Gen&Data, integrado por expertos en distintas áreas, se creó con el objetivo de analizar, desde diferentes perspectivas, la situación actual, las necesidades, los intereses y los conflictos que genera la gestión de datos genómicos en sus aplicaciones con finalidad clínica y de investigación. El resultado de su trabajo, compartiendo la experiencia y el conocimiento de estos especialistas, ha cristalizado en un documento que recoge una propuesta de criterios consensuados que sirva de orientación para todos los profesionales involucrados en la gestión de estos datos.

El Instituto Roche, que ha prestado su apoyo al grupo Gen&Data durante los casi dos años que se ha prolongado este ambicioso proyecto, se complace en presentar este documento, en lo que constituye una muestra más de nuestro decidido compromiso con el avance e implantación de la Medicina Personalizada en nuestro entorno.

Federico Plaza Piñol
Vicepresidente del Instituto Roche



INTRODUCCIÓN

Durante la última década han surgido nuevas técnicas de secuenciación genética, comúnmente llamadas *Next Generation Sequencing* (NGS) o secuenciación masiva. Estas tecnologías permiten determinar la secuencia de fragmentos de ADN o incluso el genoma completo de una persona, de una forma rápida y económica. Por esta razón, estas técnicas han irrumpido, tanto en la investigación genética como en la práctica asistencial hospitalaria, de una manera exponencial. Debido a esta rápida adopción de técnicas tan novedosas, se puede decir que las instituciones y los profesionales se enfrentan a retos y oportunidades sobre los que en ocasiones no han sido suficientemente informados y formados.

En efecto, los retos surgen en varios frentes: en la labor asistencial y en el desarrollo de la investigación; en el ámbito nacional e internacional; en el marco de la protección de los derechos de los pacientes o sujetos sanos y también de los investigadores; y en la organización y presupuestos de las instituciones, ya sean centros de investigación, asistenciales o mixtos.

En este contexto es habitual que surja un número importante de interrogantes y necesidades en distintos niveles y desde diferentes enfoques, que este trabajo pretende compilar, clasificar, discutir y, en ocasiones, responder para proporcionar pautas de trabajo.

Este trabajo está dirigido a todos los profesionales implicados en alguna medida con la NGS: consejeros genéticos, genetistas clínicos y moleculares y otros especialistas y clínicos, eticistas, profesionales del derecho, especialistas en protección de datos, bioinformáticos, gestores hospitalarios y expertos en nuevas tecnologías sanitarias.

El documento se ha estructurado en seis capítulos donde se explica qué son las técnicas de secuenciación masiva, cuál es su utilidad, sus limitaciones y las perspectivas de desarrollo; sobre qué material se llevan a cabo; qué datos se obtienen a partir de su aplicación; cómo se regula jurídicamente la gestión de estos datos; y qué medidas se pueden adoptar para una gestión eficaz y racional de la ingente información que se genera.

En algunos puntos se ha optado por abordar el análisis, tanto técnico como jurídico, distinguiendo si el ámbito en el que se obtienen los datos es clínico o de investigación. Esta aproximación ha facilitado un estudio más sistematizado de las cuestiones tratadas. No obstante,

se debe advertir que en la práctica, con frecuencia, existen conexiones estrechas entre estos ámbitos que se reflejan también en la obtención, almacenamiento y utilización de los datos. Esta realidad se ha tenido en cuenta a lo largo del trabajo.

Para la redacción del trabajo se ha seguido una metodología multidisciplinar, con reuniones entre los componentes del grupo que provienen de distintas disciplinas. Se han integrado así aportaciones desde diferentes experiencias, que se han discutido entre todos. A esto se ha añadido la autorizada revisión de expertos que no participaron en la redacción inicial del borrador pero cuya contribución ha marcado y mejorado el resultado final.

En definitiva, hemos tratado de poner a disposición de los implicados o interesados en el desarrollo y aplicación de la NGS un documento actual, riguroso y útil como material de consulta y también como análisis para la reflexión.

Pablo Lapunzina Badía y Pilar Nicolás Jiménez

Índice de capítulos

Autores	III
Presentación	V
Introducción	VII
01. Técnicas de secuenciación masiva: utilidad, oportunidades, limitaciones y perspectivas en su desarrollo	
Técnicas de secuenciación masiva: utilidad, oportunidades y limitaciones	1
Enfoque del uso clínico y de investigación	6
Tipos y variantes identificadas	7
Uso clínico y de investigación	8
Uso clínico	9
Uso en investigación	10
Indicaciones clínicas y compromisos con el paciente: asesoramiento y consentimiento informado	10
Tendencias futuras en secuenciación genómica	11
Resumen del capítulo	14
02. Muestras y genomas habituales susceptibles de análisis: ADN germinal o constitutivo, ADN somático (tumoral y no tumoral) y metagenoma	17
Secuenciación de genoma completo (WGS) o de ARN (ARNseq)	19
Secuenciación de exoma (WES) o de panel de genes	19
Secuenciación de genes o paneles mediante amplicones	19
Técnicas de inmunoprecipitación de cromatina (ChIP-Seq), secuenciación bisulfito y similares (ChIA-PET, 3C, etc.)	19
Metagenómica	20
Resumen del capítulo	21

03. Los diferentes datos que se obtienen y archivan	23
Niveles de información. Significado y relevancia clínica o para la investigación básica	23
Identificabilidad de los participantes	23
Datos no filtrados ni analizados	23
Datos filtrados y analizados. Niveles de análisis y tipo de variabilidad	25
Identificabilidad de los ficheros	28
Variantes con información de utilidad clínica	30
Informes clínicos	31
Interpretación de los resultados	34
Interacción con los profesionales solicitantes	37
Archivo y gestión de la información. Niveles de acceso	37
Archivo y niveles de acceso a la información en asistencia	38
Archivo en historia clínica	38
Almacenamiento de datos genómicos crudos	40
Archivos en servidores de la institución. Modelos organizativos	41
Archivo y niveles de acceso a la información en investigación	46
Papel potencial de los biobancos en la centralización, archivo y custodia de los datos	48
Resumen del capítulo	49
04. Marco jurídico	53
La identificabilidad del sujeto como criterio para la calificación jurídica de los datos	53
Los principios de la protección de los datos genéticos de carácter personal	58
Los sujetos implicados. Titular, responsable y encargado del tratamiento	59
Medidas de seguridad de los ficheros	62
Intervención de comités de ética	63
Comités de ética asistencial	63
Comités de ética en el contexto de la investigación	65
Papel de los CEI en la investigación con datos genéticos obtenidos a partir de bases de datos	68
Resumen del capítulo	70
05. Flujos de datos	73
Implicados en los flujos de datos en el contexto del diagnóstico genético	74
El paciente como titular de los datos	74
Información y consentimiento para la realización de diagnóstico genético	74
Derecho de acceso	76
Revocación del consentimiento y derecho de cancelación	77
Derecho a la información sobre los resultados y derecho a no saber	78
El paciente menor	79
Diagnóstico prenatal	80
El paciente fallecido	82
Los familiares del paciente	83

El clínico que solicita el análisis y la institución sanitaria	83
El personal del laboratorio que realiza la prueba y la institución que alberga el laboratorio. Posibilidad de solicitud de pruebas al extranjero	84
El profesional que lleva a cabo la interpretación clínica de los resultados y otros profesionales de la institución	84
Implicados en los flujos de datos en el contexto de la investigación	85
El sujeto de la investigación como titular de los datos	85
Información y consentimiento para la obtención y utilización de análisis genéticos con fines de investigación	85
Derecho de acceso	89
Derecho a la cancelación de datos	89
Derecho a la información sobre los resultados y derecho a no saber	90
El sujeto menor	90
Investigación con muestras o datos de fallecidos	92
Los familiares del sujeto de la investigación	92
El investigador principal, el equipo de investigación y su institución	93
Otras instituciones participantes. Posibilidad de proyectos internacionales	94
La gestión de las bases de datos genéticos a disposición de la comunidad científica	96
Resumen del capítulo	101
06. Eficacia y racionalidad en la gestión de las bases de datos genómicos: organización, planificación y sostenibilidad	103
Modelos organizativos	103
Necesidades de personal y formación	106
Infraestructuras	108
Impacto económico en asistencia e investigación	110
Adquisición de equipos de NGS	114
Estrategias de secuenciación y su impacto en el coste	114
Resumen del capítulo	116
Bibliografía	119
Anexos	123
Legislación	123
Modelos de hojas de información y consentimiento	125
Modelo para diagnóstico	125
Modelo Donación de muestras biológicas a biobanco	126
Modelo Donación de muestras biológicas a proyecto y colección o biobanco	129



TÉCNICAS DE SECUENCIACIÓN MASIVA: UTILIDAD, OPORTUNIDADES, LIMITACIONES Y PERSPECTIVAS EN SU DESARROLLO

TÉCNICAS DE SECUENCIACIÓN MASIVA: UTILIDAD, OPORTUNIDADES Y LIMITACIONES

La realización del proyecto genoma humano durante la década de 1990 puso de manifiesto la utilidad de las secuencias genéticas no solo para el avance de las investigaciones biológicas, sino también para su aplicación al campo de la medicina, la biotecnología o los estudios evolutivos, entre otros. En concreto, la finalización del proyecto genoma humano ha permitido identificar las mutaciones genéticas causantes de cientos de enfermedades hereditarias, así como el desarrollo de métodos de diagnóstico basados en la secuenciación de determinados genes que se han introducido en la práctica clínica¹. Algunas enfermedades genéticas mendelianas tienen con frecuencia una causa genética heterogénea; es decir, la misma enfermedad puede estar causada por un gran número de genes (p. ej., miocardiopatías, hipoacusias, retinosis pigmentaria, etc.), lo que obliga al estudio de numerosos genes y/o variantes genéticas para llegar a identificar sus causas. Las técnicas disponibles hasta hace un tiempo no permitían fácilmente esta aproximación, limitando por cuestiones de tiempo y coste el número de genes a estudiar por paciente. Sin embargo, los cambios genéticos responsables de muchos síndromes hereditarios aún se desconocen, por lo que es necesario secuenciar en dichos individuos un mayor número de genes, idealmente todos los genes o el genoma completo del paciente. De igual forma, numerosas patologías pueden producirse por mutaciones en un número amplio de genes, por lo que para realizar un diagnóstico preciso es necesario secuenciar un número de genes que puede variar entre unos pocos y más de un centenar.

En este sentido, es necesario destacar que la tecnología utilizada para la secuenciación de ADN fue desarrollada durante la década de 1970 por Frederick Sanger, y mejorada a mediados de la década de 1990 tras la introducción de la electroforesis capilar, por lo que este método clásico se conoce comúnmente como secuenciación Sanger o secuenciación capilar.

A pesar de que la secuenciación Sanger sigue siendo una técnica ampliamente utilizada en diagnóstico e investigación, posee una serie de limitaciones, principalmente económicas y de velocidad, que han fomentado la investigación en nuevas tecnologías de secuenciación. En concreto, durante los últimos 8 años han surgido nuevas técnicas de secuenciación masiva conocidas comúnmente bajo el nombre de *Next Generation Sequencing* (NGS). Este tipo de

técnicas son capaces de determinar la secuencia de fragmentos de ADN de una manera mucho más rápida y económica que la secuenciación tradicional². Un ejemplo muy ilustrativo para poder comparar ambas técnicas es el coste de secuenciar un genoma humano. Así, mientras que mediante la secuenciación Sanger este proceso es muy laborioso, requiere la participación de distintos equipos de trabajo durante varios meses, y tiene un coste aproximado de unos 10 millones de dólares por genoma, las técnicas de NGS permiten obtener la misma secuencia por alrededor de 1000 dólares, en unos pocos días, y tan solo requiere una pocas horas de trabajo de una sola persona. Estas características hacen que durante los últimos años se esté produciendo la implantación de la secuenciación NGS como método estándar para la mayor parte de los procesos de secuenciación, tanto en investigación como en clínica.

La dramática disminución de los costes asociados a la secuenciación ha abierto un amplio campo al desarrollo de investigaciones basadas en la secuenciación genómica y otras tecnologías ómicas (tabla 1.1). Así, en el terreno de la biomedicina se ha comenzado a secuenciar el genoma (WGS, *whole genome sequencing*) o exoma (WES, *whole exome sequencing*; las regiones del genoma que contienen la parte codificante de los genes) de una persona o familia para la identificación de mutaciones de enfermedades hereditarias cuya causa aún se desconoce³⁻⁵. En el caso de enfermedades no hereditarias, pero causadas por mutaciones genómicas, como el cáncer, distintos proyectos están secuenciando el genoma de los tumores para identificar las variantes genéticas que contribuyen al desarrollo tumoral^{6,7}. El conocimiento de estos genes permite identificar nuevas dianas terapéuticas para el desarrollo de nuevos tratamientos, así como una mejor estratificación de los pacientes, y decidir el tratamiento basándose en las alteraciones moleculares que provocan la transformación tumoral. El avance de estas técni-

Tabla 1.1. Definición de algunas aproximaciones “ómicas”

Concepto	Explicación
Genómica	Estudio del conjunto completo de genes de un organismo y/o sus organelas
Transcriptómica	Estudio del conjunto completo de moléculas de ARN mensajero presentes en una célula, tejido u órgano
Proteómica	Estudio de la totalidad de moléculas proteicas presentes en una célula, tejido u órgano
Epigenómica	Estudio de las complejas modificaciones que experimenta la cromatina, tanto en su modificación química como en los cambios topológicos condicionados por efectos internos y ambientales, y que incluye todos aquellos procesos que alteran la expresión de genes sin cambiar la secuencia del ADN
Farmacogenómica	Estudio de cómo la información genética de una persona afecta a la respuesta del organismo a un fármaco, eminentemente en términos de toxicidad y/o eficacia de la respuesta terapéutica
Filogenómica	Estudio mediante la comparación de secuencias de genoma completo de individuos de la escala evolutiva para reconstruir la historia evolutiva de los organismos
Metabolómica	Estudio del conjunto completo de metabolitos (intermediarios de bajo peso molecular) en una célula, tejido u órgano
Metagenómica	Estudio de los organismos que cohabitan con otros organismos (en general microbios que cohabitan con vertebrados) y que enfoca más allá del organismo individual para centrarse en los genes comunitarios y cómo pueden estos influir en las actividades del otro en servicio a las funciones colectivas
Fenómica	Estudio de todos los hallazgos fenotípicos (conjunto de signos y síntomas) de un organismo en un momento dado

cas está también permitiendo utilizar la secuenciación NGS para la detección de enfermedad mínima residual en cáncer, o para diagnóstico prenatal no invasivo mediante la secuenciación del ADN fetal circulante en sangre materna⁸. La secuenciación del genoma de una persona también permite identificar todas las variantes genéticas que pueden influir en el metabolismo de fármacos, por lo que podría acelerar la investigación en farmacogenómica, limitada en la actualidad por el elevado coste de la secuenciación de todos los genes implicados en la respuesta a los fármacos.

Gracias a estas nuevas tecnologías, ahora es posible secuenciar el genoma o exoma de decenas de miles de individuos a nivel mundial, y con el continuo abaratamiento es probable que contribuya a su implantación gradual no solo en proyectos de investigación, sino también en la práctica clínica.

El campo de la oncología ha sido uno de los más beneficiados por estos avances técnicos, cuya introducción ha supuesto una verdadera revolución del estudio de la biología tumoral. La expresión “genómica del cáncer” alude al estudio de genomas tumorales a diferentes niveles, entre ellos, cambios en la secuencia de ADN (alteraciones del número de copias, mutaciones y reordenamientos), el epigenoma (patrones de metilación del ADN y de modificación de histonas) y el transcriptoma (alteraciones en la expresión de genes o microARN) (tabla 1.1)⁹. La NGS permite identificar tanto alteraciones que desempeñan funciones esenciales en la transformación y progresión neoplásica (mutaciones “drivers” o conductoras) como alteraciones no fundamentales concomitantes conocidas como mutaciones “passenger”. La identificación de genes y vías importantes desde el punto de vista biológico en diferentes tipos de cáncer puede generar información diagnóstica, pronóstica y terapéutica clínicamente relevante. También es posible que un subgrupo de mutaciones en oncogenes o en otras regiones genómicas sea susceptible de actuación farmacológica (*druggable*), es decir, que se identifiquen como dianas para aplicar o desarrollar tratamientos¹⁰. Por todas estas razones, las investigaciones actuales están provocando una reclasificación de los tumores en función de criterios genéticos.

Los avances en el conocimiento de la genética del cáncer permitieron identificar genes de alto riesgo de padecer diferentes tipos de cáncer, entre los que se encuentran los tumores más prevalentes, como el cáncer colorrectal y el cáncer de mama. Las alteraciones en la línea germinal de estos genes son poco frecuentes en la población, pero confieren un elevado riesgo relativo de padecer cáncer a los portadores de dichas alteraciones.

Los avances científicos y tecnológicos están revolucionando la aproximación a la evaluación del riesgo genético de padecer muchas enfermedades y cáncer, a su detección y prevención y a la terapia dirigida, cumpliendo con la promesa de la medicina personalizada.

Por lo que se refiere a la calidad de los resultados, como se ha comentado anteriormente, la obtención de datos genéticos a partir de secuenciación NGS implica varias etapas y métodos de análisis, cada uno de los cuales puede influir en la calidad de los resultados que se generen (fig. 1.1). Los factores que influyen en la calidad de los resultados obtenidos son muy variados: conservación de la muestra (tejido fresco o fijado), pureza de la muestra (en el caso de tumores, porcentaje de células no tumorales presentes en la muestra), técnica de secuenciación, cobertura, herramientas de análisis, etc. Cada uno de estos factores va a influir en el tipo de errores cometidos y en la fiabilidad de los datos derivados del análisis.

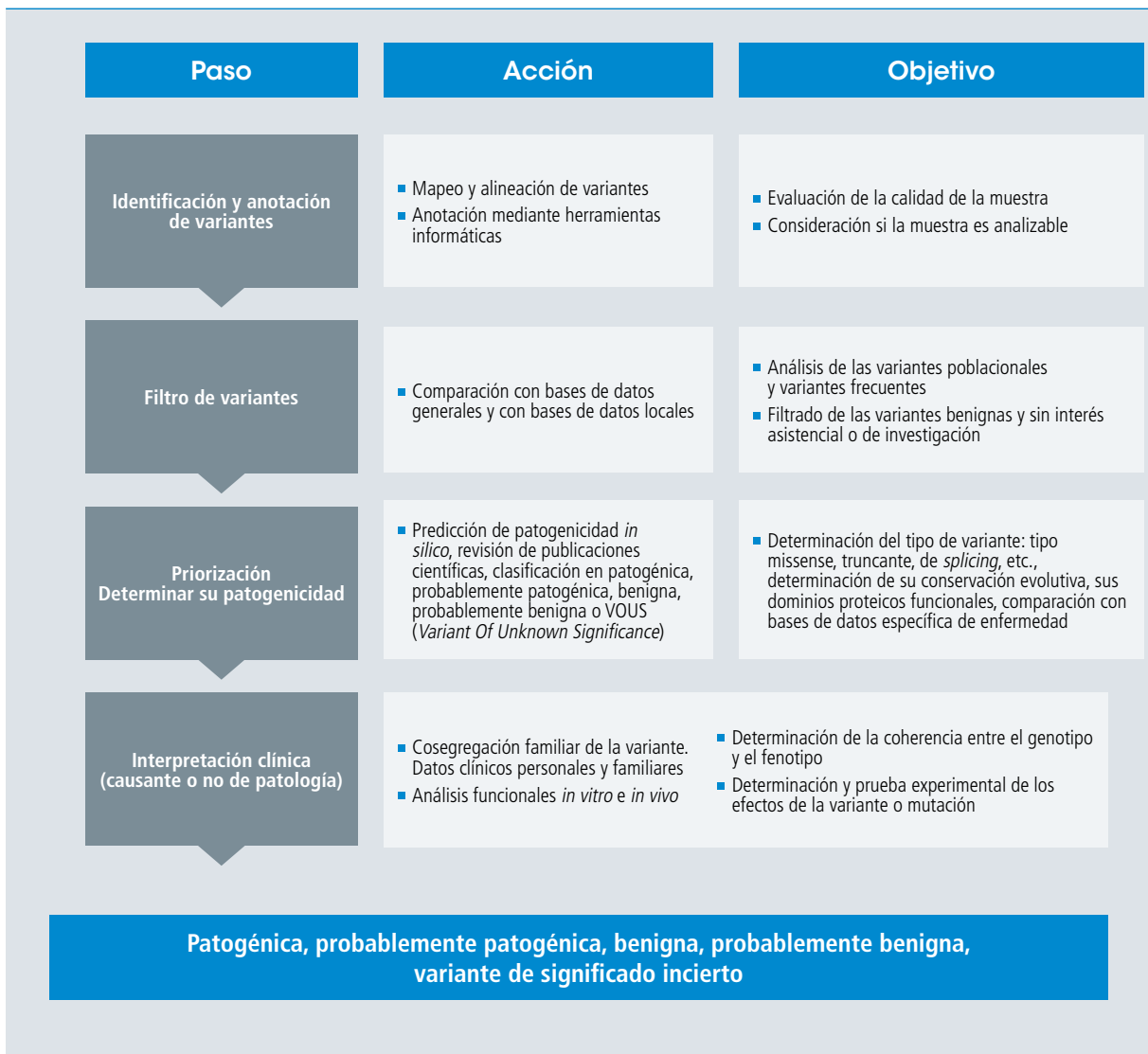


Figura 1.1. Algoritmo o *pipeline* clásico de análisis de estudios de NGS.

Así, en el caso de los datos obtenidos para investigación básica, cuyo objetivo no es el diagnóstico de un paciente, sino una investigación que pueda contribuir a profundizar en el conocimiento de una patología, el investigador puede decidir aceptar tasas de error superiores a las que se permitirían en el caso de pruebas diagnósticas. Estas últimas deben ser realizadas por laboratorios acreditados para ello, y con unos controles estrictos que permitan disminuir la posibilidad de error en el diagnóstico. En cualquier caso, se debería proporcionar una serie de parámetros estadísticos que reflejasen las limitaciones de los análisis genéticos, para así poder tomar decisiones informadas por parte de los profesionales médicos y pacientes. A continuación se definen los principales parámetros necesarios para interpretar este tipo de resultados (tabla 1.2).

Tabla 1.2. Conceptos de cobertura, profundidad, sensibilidad y especificidad en los estudios de NGS

Concepto	Explicación
Cobertura	Cantidad (porcentaje) de secuencia de interés de un gen, conjunto (paneles) o de la totalidad de las regiones codificantes de los genes (exoma) que está representado en la región de captura o región a analizar. Las técnicas de captura o de amplificación de las secuencias, así como la propia secuenciación, pueden provocar que alguna región concreta no se secuencia completamente y por ello se pierda información de una o varias regiones. Se debe establecer qué porcentaje de cobertura mínimo se necesita para dar un diagnóstico, y en caso contrario repetir una prueba. Es aconsejable hacer constar en el informe este parámetro.
Profundidad	Número de veces que cada base del genoma está presente en los <i>reads</i> (lecturas) de secuenciación producidos. Este valor se denomina profundidad de cobertura (<i>depth of coverage</i>, o simplemente, <i>coverage</i>) y es uno de los factores determinantes para evaluar la fiabilidad del nucleótido asignado a esa posición del genoma. La profundidad de secuencia, o algunas veces llamada profundidad de cobertura, se refiere al número de veces que se ha leído la misma región del genoma. Así, en los protocolos de secuenciación es habitual generar una cantidad de datos que permiten leer cada base de un gen entre 30 y 100 veces, de tal forma que si hay una mutación, varias de esas lecturas contendrán la misma mutación, a diferencia de si hay un error de secuenciación, en cuyo caso la probabilidad de que haya más de una lectura con el mismo error es muy baja.
Sensibilidad	Capacidad de una técnica para identificar variantes/mutaciones que están presentes en la muestra, de tal forma que si un paciente tiene una mutación, la sensibilidad refleja cuál es la capacidad de la técnica para detectarla. Aunque idealmente sería deseable tener una sensibilidad del 100%, en la práctica la sensibilidad puede ser menor. Influyen en la sensibilidad una cobertura o una profundidad insuficientes, así como el gen y la mutación en cuestión o su presencia en la totalidad de las células o en mosaicismo (tumor) y las características de la muestra en estudio (es muy frecuente que una muestra tumoral no solo tenga células tumorales, sino también células no tumorales, lo que dificulta la identificación de mutaciones en el tumor). Es necesario establecer <i>a priori</i> la sensibilidad de una técnica para distintos tipos de pruebas, utilizando el dato del porcentaje de falsos negativos para muestras con variantes/mutaciones conocidas, así como reflejar dicho parámetro en los informes correspondientes, ya que puede ser de utilidad para la interpretación clínica de los resultados obtenidos.
Especificidad	La especificidad refleja la fiabilidad de las variantes identificadas y se define como el porcentaje de variantes génicas que la técnica detecta y que están realmente presentes en la muestra (y no son falsos positivos). Se puede calcular verificando, mediante una técnica de secuenciación ortogonal (distinta de la utilizada inicialmente), las distintas mutaciones encontradas, y calcular qué porcentaje de ellas se confirman. Al igual que en los casos anteriores, es deseable que la especificidad de una prueba diagnóstica sea lo más próximo al 100% posible, lo que minimiza el número de falsos positivos. Dependiendo del tipo de prueba y patología estudiada se puede estimar la especificidad de una prueba, y en cualquier caso, una mutación concreta puede ser posteriormente verificada mediante la realización de una segunda prueba de secuenciación con una tecnología distinta. Y el dato debe reflejarse en los informes. Dado que tanto la sensibilidad como la especificidad no son homogéneas, sino que dependiendo del tipo de variante genética estudiada varía (sustituciones, pequeñas inserciones/deleciones, variaciones estructurales), la capacidad de detección es muy distinta, por lo que se debería incluir esta información detallada para conocer, para cada tipo de variación, qué fiabilidad tiene.

ENFOQUE DEL USO CLÍNICO Y DE INVESTIGACIÓN

El uso de la NGS ha permitido en los últimos años un desarrollo muy importante del diagnóstico genético de laboratorio, así como del conocimiento generado por la investigación sobre las bases genéticas de muchas enfermedades. Con frecuencia, la distinción entre diagnóstico e investigación es difícil y debe definirse con precisión pues implica compromisos muy diferentes entre el médico y el paciente, y el investigador en su caso. A continuación detallamos los aspectos fundamentales del uso de la NGS para la asistencia clínica y para la investigación, que se reflejan en la tabla 1.3 de modo resumido. Debemos primero establecer la distinción entre las diferentes aproximaciones tecnológicas al uso de la NGS.

El diseño de paneles de NGS establece la captura de exones y regiones intrónicas flanqueantes de un número de genes determinado, debe garantizar una adecuada profundidad de la cobertura en todos los genes analizados y generalmente se acompaña de la pirosecuenciación o de cualquier otra técnica complementaria que pueda confirmar o descartar los hallazgos en las regiones conocidas con cobertura insuficiente.

La secuenciación de todo el exoma (WES) aborda la captura de todos los exones conocidos así como de regiones flanqueantes exón-intrón potencialmente implicadas en el *splicing*. Se

Tabla 1.3. Uso clínico y de investigación de las técnicas de NGS

	Indicaciones	Tipo de técnica	Asesoramiento genético	Información de resultados	Custodia de los datos y acceso futuro
Uso clínico de la NGS	■ Necesidad diagnóstica o reproductiva ■ Opciones terapéuticas o reproductivas	■ Buena capacidad de detección de la causa del fenotipo ■ Poca información incierta o incidental ■ Paneles y WES, no WGS por ahora	■ Basado en las necesidades médicas y reproductivas ■ Necesidad de informar hallazgos incidentales graves e intervenibles ■ Conveniencia de informar a otros familiares	■ Tiempo de respuesta adecuado ■ Valoración de la certeza de los resultados antes de su información ■ Su uso reproductivo debe adecuarse a la ley y la ética	■ Custodia de la secuencia ■ Custodia de la confidencialidad ■ Mecanismos para acceder a los datos y a revisar su interpretación ■ Acuerdos entre las instituciones y centros, de acuerdo a la ley
Uso de investigación de la NGS	■ Tipo de proyecto y financiación ■ Expectativas de resultados	■ Máxima capacidad de detección de nuevos hallazgos ■ El estudio de múltiples familiares puede ser necesario ■ WES y WGS, uso limitado de paneles	■ El asesoramiento en investigación debe proporcionarse con los resultados genéticos positivos, por un investigador clínico con formación suficiente ■ Necesidad de informar hallazgos causales de la enfermedad y hallazgos incidentales graves e intervenibles	■ Información de resultados definitivos si son clínicamente relevantes ■ Los usos futuros de los hallazgos pasan por su confirmación clínica y cumplir los requisitos de uso clínico	■ Custodia de los datos y mecanismos de acceso futuro ■ Pactos particulares para cada proyecto de investigación de acuerdo a la ley

WES: whole exome sequencing; WGS: whole genome sequencing.

calcula que la captura de este 2% del genoma podría detectar aproximadamente el 80-85% de los cambios de secuencia responsables de las enfermedades mendelianas. Se reservan las comprobaciones para los genes definidos como de interés.

La secuenciación del genoma (WGS), pudiendo detectar cambios en regiones no codificantes también, ofrece la posibilidad de detectar las mutaciones responsables del 20% de enfermedades mendelianas restantes. Ofrece además la posibilidad de conocer alteraciones en secuencias reguladoras, de identificar múltiples variantes implicadas en la variabilidad fenotípica dentro la normalidad o en el modelo predominante multifactorial de las enfermedades comunes y, por tanto, en la variabilidad de la penetrancia y la expresión que determinan el fenotipo de los afectados por las enfermedades mendelianas. Pero el número de variantes inciertas es muchísimo mayor en el WGS que en los estudios de WES, que a su vez es mayor que con el uso de paneles.

Otros estudios encaminados a determinar las consecuencias reales de las mutaciones en el ADN genómico sobre el ARN, como el ARN-Seq y otros estudios de transcriptómica y proteómica, permiten conocer el impacto transcripcional y bioquímico de las mutaciones y, por tanto, acercarse a la patogenia y fisiopatología de las enfermedades genéticas.

Debemos definir brevemente el tipo de relevancia clínica asociada a los diferentes tipos de variantes de secuencia identificadas mediante las diferentes técnicas de NGS, ya que el uso idóneo de estas técnicas, su interpretación adecuada y su reflejo en los informes se tratan con mayor detalle en las páginas 30 y 31 de este documento.

Tipos de variantes identificadas

La variabilidad de la secuencia del genoma humano es enorme, pero el conocimiento de su significado es aún muy incompleto. El conocimiento es mayor cuando las variantes ocurren en genes previamente implicados en enfermedades humanas, principalmente mendelianas. Para estas secuencias principalmente exónicas se ha generado ya un gran conocimiento del significado de los cambios, que permite asignar el valor preciso de muchas de las variantes en relación con la enfermedad de estudio, tanto en población normal como en población afectada, en las correspondientes bases de datos. Sin embargo, el conocimiento del significado de las variantes del ADN genómico en su conjunto, incluyendo regiones intrónicas profundas y regiones sin genes que constituyen la gran mayoría del genoma, es mucho más limitado.

Las variantes de secuencia pueden ser benignas, patogénicas o inciertas. Para definir que una variante es benigna debe haber constancia de que su presencia nunca causa un fenotipo anómalo. Para genes bien conocidos, existen ya datos de variación humana que permiten definir su frecuencia en la población normal. Los estudios actuales con frecuencia no permiten aún descartar la intervención de estas variantes benignas en un modelo multifactorial de enfermedad, y que estas determinen una cierta predisposición a enfermedades o defectos congénitos, pero al menos se ha descartado con evidencia razonable su intervención en la enfermedad mendeliana.

Una variante es patogénica cuando existen datos suficientes para implicarla causalmente en un fenotipo anómalo. Si el fenotipo asociado a la variante es el interrogado por una prueba

clínica o el investigado en su caso, se trata de un hallazgo causal de la enfermedad. Si el fenotipo asociado a la variante es otro, no relacionado con la enfermedad de estudio, se trata de un hallazgo patogénico incidental. Los hallazgos patogénicos incidentales pueden representar la predicción de enfermedades incluso más relevantes para la salud del paciente que el fenotipo en estudio, por lo que la transmisión al paciente de estos riesgos debe ser meditada y contemplada particularmente en el asesoramiento genético y en el consentimiento informado.

De entre las variantes patogénicas halladas en un estudio, cada vez es más frecuente el hallazgo de variantes de penetrancia incompleta o susceptibilidad, que intervienen en el riesgo para presentar un fenotipo pero no lo determinan ni cuantitativamente (penetrancia incompleta, a veces baja), ni cualitativamente (fenotipo muy variable en su expresión). La variabilidad en la expresión se refleja particularmente en la asociación de una variante (CNV o número de copia, o bien variante de secuencia en un gen concreto) a fenotipos neurogenéticos que incluyen epilepsia, trastornos del espectro autista (TEA), dificultades de aprendizaje y discapacidad intelectual (DI), incluso enfermedades psiquiátricas del adulto como la esquizofrenia. Estas variantes actúan claramente en un modelo poligénico y multifactorial, pero hoy apenas conocemos una pieza del modelo.

Desafortunadamente, muchas variantes no pueden ser clasificadas y permanecen como variantes inciertas (*variants of unknown significance*: VUS o VOUS). Más adelante, en este documento se discute ampliamente sobre la asignación de certeza a una variante, y hay que entender que su relevancia es mucho mayor para el uso clínico de la técnica que para el uso de investigación. Una variante incierta puede ser probablemente patogénica en el caso de que haya precedentes en modelos animales que impliquen al gen, que las funciones de este y sus vías de señalización parezcan implicadas y, solo en estos casos, se deberá proceder a la información al paciente y la segregación de la variante con el fenotipo. Demostrar que se trata de una variante *de novo* en un caso de fenotipo *de novo* o que esta variante se repite en fenotipos similares o asociados en otros miembros de la familia puede facilitar el reconocimiento de su patogenidad. Por ello, la disponibilidad de muestras familiares y la información del proceso a los familiares implicados es de especial relevancia.

Uso clínico y de investigación

Dos factores determinan la elección de la técnica de NGS (paneles, WES, WGS) a elegir: la tasa de detección potencial de la técnica para el fenotipo interrogado, y desde luego también la frecuencia de hallazgos inciertos o incidentales poco relevantes de estos.

Los paneles de NGS se diseñan para estudiar uno o varios fenotipos asociados de alguna forma (p. ej., displasias esqueléticas, o bien DI), por lo que la capacidad de hallar la variante patogénica depende de cuánto conocemos los genes asociados a este fenotipo, y puede ser elevada si el conocimiento es amplio. Lo es para las displasias esqueléticas, pero no lo es para la DI, ya que aún se conoce solo una proporción pequeña de los genes causantes del fenotipo. Pero, para los paneles, la capacidad de generar información sobre variantes incidentales e inciertas es muy limitada. En cambio, las técnicas de WES y WGS ofrecen la posibilidad de identificar nuevos genes asociados al fenotipo, pero incrementan mucho la posibilidad de información incidental y de significado incierto.

Evidentemente, los costes de los estudios y la disponibilidad técnica y bioinformática en los diferentes centros son factores importantes también para su elección. Tanto paneles de NGS como estudios de WES tienen en 2015 precios razonables, entre los 1000 y los 2000 €, y la elección de uno u otro depende más del fenotipo de estudio y la capacidad de detección prevista que del coste. El estudio de WGS es aún muy caro para su uso clínico sistemático y también en investigación, aunque esto tenderá a cambiar.

Hay que tener en cuenta que el uso inicial de la técnica para investigación puede generar un resultado de investigación con significado clínicamente relevante y debe ser transmitido por un clínico o un asesor genético, debe ser confirmado por un laboratorio clínico, y a partir de ese momento se aplican los requisitos para el uso clínico y reproductivo de esta información.

USO CLÍNICO

La elección para el diagnóstico clínico de una de las técnicas de NGS debe estar determinada por la mayor capacidad de detección, el menor coste, el menor tiempo de diagnóstico y la menor información incidental e incierta posibles. Esto se hace aún más importante en diagnóstico prenatal, que puede llevar a opciones reproductivas como la terminación de la gestación, con implicaciones éticas y legales inmediatas. Lo mismo ocurre con el diagnóstico preimplantacional.

Por tanto, cuando existan paneles de secuenciación para determinadas patologías, el conocimiento de los genes implicados sea amplio y bueno, particularmente del significado de la variación de secuencia de los genes (bases de datos), y se garantice cobertura y profundidad, deberán elegirse estos. El uso del WES en estas circunstancias también es adecuado si la interpretación bioinformática es experta y la cobertura y profundidad son buenas o se completan mediante otras técnicas para los genes conocidos implicados en el fenotipo.

Cuando solo se conoce una parte de los genes implicados en el fenotipo, o el fenotipo observado es poco específico, el WES ofrece mayor capacidad de detección a un buen coste. El conocimiento de los genes potencialmente implicados en el fenotipo (conocimiento de las vías de señalización, modelos animales o celulares, genes de fenotipos humanos similares), y el conocimiento del patrón de herencia (autosómica dominante, recesiva con o sin consanguinidad, ligada al cromosoma X) puede llegar a predecir una posible homocigosidad o a sugerir la localización del gen no conocido. Son estos casos en los que el WES tendrá mayor capacidad de detección. Hay que tener en cuenta que fenotipos como la DI o el TEA tienen como base mutaciones autosómicas dominantes *de novo*, y solo el WES/WGS del trío (padre, madre y caso índice) podrá ayudar en el proceso. También enfermedades dominantes o recesivas pueden requerir el estudio completo mediante WES o secuenciación variante de varios familiares para asignar patogenicidad cierta a una de las variantes encontradas, y en todos estos casos los costes son más elevados, pero si no se hace así el estudio no será útil.

Por supuesto, la información incierta e incidental potencialmente generada es mucho mayor, por lo que los compromisos adquiridos con el paciente mediante el asesoramiento genético y el consentimiento informado son elementos importantes, aún más cuando se plantee el uso clínico de WGS.

USO EN INVESTIGACIÓN

El uso en investigación de los paneles de NGS se fundamenta con frecuencia en estudiar su capacidad de detección para el fenotipo interrogado. Puede también ser útil para complementar investigaciones más amplias con un cribado genético razonable por su capacidad de detección, su coste y su interpretación más fácil que otras técnicas. Pero es obvio que los estudios de WES y WGS ofrecen muchas más posibilidades de hallazgos, como nuevos genes, la implicación de varios genes en el fenotipo que nos acerca al conocimiento del modelo multifactorial de muchas enfermedades, y para ayudar así a resolver los enigmas planteados por la penetrancia incompleta y la variabilidad en la expresión.

Los objetivos de la investigación determinarán cuántos pacientes de la misma familia o cuántas familias con fenotipo similar deberán estudiarse para poder encontrar la base genética de la enfermedad y también las técnicas a utilizar.

La información incidental generada y la información incierta son aquí mucho más grandes, pero los compromisos del investigador con el paciente pueden definirse mucho mejor de acuerdo a cada proyecto y circunstancias, respetando la ley y las recomendaciones actuales sobre el manejo de la información.

Así como el uso clínico de las técnicas debe contar con un coste y un tiempo de respuesta bien definido y razonable, y es el médico responsable o el asesor genético quienes transmitirán los resultados, todos estos factores deben definirse con precisión en la investigación de manera mucho más abierta. Insistimos en que un resultado de investigación relevante deberá entrar en el circuito de uso clínico y tratarse como tal para el paciente y su familia.

Indicaciones clínicas y compromisos con el paciente: asesoramiento y consentimiento informado

Los compromisos con el paciente por parte del médico que le atiende o por parte del investigador se establecen a través del proceso de información y del consentimiento.

El asesoramiento previo a la prueba genética es absolutamente necesario con estas técnicas. De hecho, la atención médica junto con el consejo genético determina si estamos en un marco de investigación o de asistencia. Cuando una enfermedad tiene una gravedad significativa, la familia tiene un riesgo elevado, y manifiesta su deseo de tener opciones reproductivas para evitar nuevos casos, deberemos ofrecer el uso clínico de estas pruebas para diagnosticar al caso índice y después establecer el riesgo de los familiares. Evidentemente, para la valoración sobre si esta enfermedad es susceptible de diagnóstico prenatal e interrupción del embarazo o bien de selección embrionaria por desarrollar diagnóstico genético preimplantacional (DGP), deberemos adecuarnos a la ética y la ley para determinar, antes de hacer el estudio, cuál va a ser su significado para la familia y las opciones que se derivan. Es claro que cualquier paciente con una enfermedad genética merita un diagnóstico preciso, y ahora o más adelante el conocimiento de la base molecular puede ayudar para nuevos tratamientos y el seguimiento de los pacientes en función de sus riesgos. Pero de manera inmediata, teniendo en cuenta los costes y la disponibilidad de la tecnología, el centro y el médico deberán establecer su orden de prioridades para ofrecer estos estudios a los pacientes y pactar bien el uso futuro de la

información. También antes de la prueba deberemos informar de la posibilidad de variantes inciertas y de hallazgos incidentales. Algunos de estos hallazgos incidentales pueden tener grandes repercusiones para la salud¹¹, y deberán informarse siempre, algo que también deberá saber el paciente antes de la prueba. Esta lista irá variando, pero el criterio de utilidad clínica del conocimiento de la gravedad de las repercusiones de salud de una mutación deberá prevalecer como criterio principal para el asesoramiento genético y reproductivo. La necesidad de pactar con el paciente determinadas condiciones para la información (no conocer los hallazgos inciertos, ni las variantes de bajo o moderado riesgo para padecer una enfermedad) puede reflejarse en el consentimiento informado, que deberá contener muchos de los elementos expuestos en el asesoramiento previo a la prueba genética.

El clínico o el asesor genético adquieren el compromiso de gestionar los resultados del estudio, a través del asesoramiento posterior al estudio genético y también mediante un programa de actuación médica concreto. Se informará de los resultados patogénicos principalmente y de aquella información incierta e incidental que se considere relevante. Como en el resto de pruebas genéticas, pero aún más por el número de variantes encontradas, es necesario con frecuencia segregar las variantes con los progenitores y a veces con otros familiares. Esto también debería exponerse antes de hacer la prueba, y quizá informar a los implicados, ya que estas comprobaciones pueden ser imprescindibles para un resultado útil del estudio para el paciente y su familia. Esta necesidad es tan frecuente que muchos centros realizan ya la extracción de los progenitores o de casos índice adicionales a la vez que al paciente en estudio.

El consentimiento informado debe también ajustarse a normas establecidas, y el paciente debe tener el derecho de revocación. La posibilidad de dar opciones a los pacientes para decidir si desean o no conocer ciertos tipos de información debe contemplarse.

Los compromisos del investigador con el paciente y su médico pueden variar mucho más, pero a la vez deben definirse con mayor precisión en el asesoramiento y en el consentimiento. A los pacientes que participan en la investigación siempre hay que devolverles el conocimiento de los hallazgos patogénicos causantes de su fenotipo y de los hallazgos incidentales importantes. Pero, en cambio, el resto de hallazgos pueden no comunicarse nunca.

En ambos casos, pero aún más en la investigación pues trasciende con frecuencia el ámbito sanitario, debe garantizarse la custodia de los datos y la confidencialidad de manera exquisita como se explica en otras partes del documento.

Un hecho importante es que las secuencias generadas por NGS, sobre todo por WES o WGS, siempre son interpretadas a la luz del conocimiento en el momento del estudio. Pero la información sobre el significado de las variantes de secuencia va a ir aumentando constantemente, por lo que la custodia de la secuencia debe garantizarse y establecerse de manera organizada en todos los casos, así como el acceso a estos para nuevos análisis.

TENDENCIAS FUTURAS EN SECUENCIACIÓN GENÓMICA

De cara al futuro es ciertamente previsible que la secuenciación vaya a estar presente en todos los ámbitos de la sanidad y en muchos otros de la vida de los ciudadanos. La rápida evolución

de la tecnología de secuenciación hace previsible la generalización de nuevos métodos de secuenciación que incluirán la obtención de secuencias completas de grandes fragmentos de ADN –posiblemente cromosomas completos– de pequeños grupos de células o incluso células individuales. Los primeros prototipos operativos de este tipo de secuenciadores ya están disponibles^{12,13}.

Estas mejoras y extensiones de las tecnologías existentes supondrán un cambio radical en los protocolos experimentales asociados. Así, mientras que en la actualidad lo habitual es secuenciar una única muestra de un paciente, y obtener muchas secuencias de la misma región (o de todo el genoma) para tener suficiente cobertura para identificar variantes génicas, en el futuro es posible que las repeticiones técnicas dejen de tener sentido, mientras que las repeticiones biológicas adquirirán un significado distinto, y podrán referirse a la secuenciación de numerosas moléculas de ADN de distintas células de una misma muestra, repeticiones cuya principal limitación será el origen biológico de las muestras¹⁴.

A su vez, es más que previsible que junto a la secuenciación de ADN se generalizará la secuenciación de ARN (ARN-seq), lo que permitirá monitorizar de modo directo y continuo estados fisiológicos y celulares. A diferencia de los arrays de expresión¹⁵ ya empleados en clínica, el ARN-seq no solo permite cuantificar la expresión de todos los genes de una muestra, sino que permite identificar cambios en la expresión de isoformas debido a fenómenos de *splicing* alternativo, o incluso nuevos fenómenos de *splicing*, así como detectar la existencia de genes de fusión, expresión específica de alelo, etc., que no pueden ser detectados mediante el análisis de arrays de expresión. Mientras que las primeras aplicaciones en ARN-seq están disponibles y comienzan a ser evaluadas, es evidente que todo este tema debe pasar todavía por un proceso de regularización y estandarización de la metodología.

Los cambios en los protocolos experimentales y en el tipo de resultados que se obtendrán en secuenciación tanto de ADN como de ARN requerirán el desarrollo paralelo de nuevas aproximaciones tanto para la gestión y el almacenamiento de datos como para estimar la calidad y fiabilidad de los resultados, que afectarán a la propia interpretación científica de los mismos. Por ejemplo, los problemas actuales para el almacenamiento de grandes cantidades de datos de secuenciación para cada muestra –asociado a las tecnologías de secuenciación de fragmentos cortos– serán reemplazados por problemas de almacenamiento de fragmentos largos secuenciados repetidamente. Es, por tanto, importante prever la evolución desde la situación técnica actual, completamente dependiente de una tecnología dominante de secuenciación de ADN, hacia los retos que los cambios tecnológicos implicarán.

De cara al futuro es también relevante resaltar la importancia que muy pronto adquirirán las diferencias entre poblaciones humanas en el proceso de análisis de los datos genómicos. En la situación actual, las operaciones de almacenamiento y análisis se realizan considerando una versión única del genoma humano como referencia, es decir las mutaciones se refieren a unas coordenadas proporcionadas por una versión concreta del ensamblado del genoma humano –o de las referencias correspondientes en el caso de otras especies–. A medida que el conocimiento sobre las diferencias entre individuos avanza, esta aproximación se hace poco sostenible, particularmente para el análisis de las variantes individuales en secuencias repetidas (CNV) y otras variaciones estructurales relacionadas con reorganizaciones locales de regiones del genoma¹⁶. De hecho, el uso de una única referencia es potencialmente erróneo, puesto

que puede confundir variantes específicas comunes en determinadas poblaciones con variantes individuales que en otras poblaciones pueden estar asociadas a riesgo de una determinada enfermedad. Por estas razones, muy pronto se comenzarán a usar genomas de referencias específicos de poblaciones como una primera aproximación, y en un futuro no muy distante se considerará cualquier genoma en su contexto, tanto individual como poblacional. En este caso, será necesario implementar sistemas que regeneren la secuencia completa de genomas individuales (ensamblado de genomas de nuevo), y cambiar el modo en que se almacena y anota la información genómica, y por supuesto se requerirán nuevos sistemas de representación e interpretación de la información genómica individual.

Finalmente, mientras que el énfasis actual, particularmente en el ámbito más traslacional, se refiere básicamente a la secuenciación de ADN y ARN, es evidente que otros tipos de información son igualmente necesarios para poder extrapolar entre el nivel molecular y el fenotípico, enfermedad y estados fisiológicos. Este tipo de información adicional incluye, entre otros: datos epigenéticos, tanto secuenciación de modificaciones de las bases del ADN (methyl-seq) como secuenciación de elementos de control génico (chip-seq) como sitios de unión de factores de transcripción, regiones de unión de factores de modificación epigenético u otros factores que pueden modificar el nivel de expresión génica; datos de proteómica y metabolómica, cuya obtención es todavía demasiado lenta y costosa, pero que sin duda serán necesarios en el contexto general del análisis de las bases moleculares de las enfermedades, y además son potencialmente capaces de proveernos de mejores marcadores de enfermedades que los resultantes de la secuenciación de ADN.

En definitiva, la explosión técnica actual en genómica debe verse a la luz del abismo existente entre información molecular y fisiológica; es importante considerar críticamente las limitaciones que la tecnología actual de secuenciación impone y estar preparado para los cambios que la muy rápida evolución de la tecnología va a producir. Sea cual sea el devenir de la tecnologías genómicas, el problema de interpretación de los resultados permanecerá como el mayor reto tanto científico como traslacional.

RESUMEN DEL CAPÍTULO

La finalización del proyecto genoma humano ha permitido identificar las mutaciones genéticas causantes de miles de enfermedades hereditarias, así como el desarrollo de métodos de diagnóstico basados en la secuenciación de determinados genes que se han introducido en la práctica clínica. Las técnicas disponibles hasta hace un tiempo no permitían fácilmente esta aproximación, y limitaban por cuestiones de tiempo y coste el número de genes a estudiar por paciente. Durante los últimos años han surgido nuevas técnicas de secuenciación masiva conocidas comúnmente bajo el nombre de *Next Generation Sequencing* (NGS). Este tipo de técnicas son capaces de determinar la secuencia de fragmentos de ADN de una manera mucho más rápida y económica que la secuenciación tradicional. La notable disminución de los costes asociados a la secuenciación ha abierto un amplio campo al desarrollo de investigaciones basadas en la secuenciación genómica y otras tecnologías ómicas.

En el terreno de la biomedicina se ha comenzado a secuenciar el genoma (WGS, *whole genome sequencing*) o exoma (WES, *whole exome sequencing*; las regiones del genoma que contienen la parte codificante de los genes) de una persona o familia para la identificación de mutaciones de enfermedades hereditarias cuya causa aún se desconoce. El avance de estas técnicas está también permitiendo utilizar la secuenciación NGS para la detección de mutaciones responsables de neoplasias o de enfermedad mínima residual en cáncer, o para diagnóstico prenatal no invasivo mediante la secuenciación del ADN fetal circulante en sangre materna. La secuenciación del genoma de una persona también permite identificar todas las variantes genéticas que pueden influir en el metabolismo de fármacos, por lo que podría acelerar la investigación en farmacogenómica.

La variabilidad de la secuencia del genoma humano es enorme, pero el conocimiento de su significado es aún muy incompleto. De forma genérica, las variantes de secuencia pueden ser benignas, patogénicas o inciertas. De entre las variantes patogénicas halladas en un estudio, cada vez es más frecuente el hallazgo de variantes de penetrancia incompleta o susceptibilidad, que intervienen en el riesgo para presentar un fenotipo, pero no lo determinan ni cuantitativamente (penetrancia incompleta, a veces baja), ni cualitativamente (fenotipo muy variable en su expresión).

La obtención de datos genéticos a partir de secuenciación NGS implica varias etapas y métodos de análisis, cada uno de los cuales puede influir en la calidad de los resultados que se generen. Los factores que influyen en la calidad de los resultados obtenidos son muy variados: conservación de la muestra (tejido fresco o fijado), pureza de la muestra (en el caso de tumores, porcentaje de células no tumorales presentes en la muestra), técnica de secuenciación, cobertura, herramientas de análisis, etc.

Existen diferentes aproximaciones tecnológicas con el uso de la NGS. El diseño de paneles de NGS establece la captura de exones y regiones intrónicas flanqueantes de un número de genes determinado. La secuenciación de todo el exoma (WES) aborda la captura de todos los exones conocidos así como de regiones flanqueantes exón-intrón potencialmente implicadas en el *splicing*. Se calcula que la captura de este 2 % del genoma podría detectar aproximadamente el 80-85 % los cambios de secuencia responsables de las enfermedades mendelianas. La secuenciación del genoma (WGS), pudiendo detectar cambios en regiones no codificantes tam-

bién, ofrece la posibilidad de detectar las mutaciones responsables del 20% de enfermedades mendelianas restantes. Ofrece, además, la posibilidad de conocer alteraciones en secuencias reguladoras, de identificar múltiples variantes implicadas en la variabilidad fenotípica dentro de la normalidad o en el modelo predominante multifactorial de las enfermedades comunes y, por tanto, en la variabilidad de la penetrancia y la expresión que determinan el fenotipo de los afectados por las enfermedades mendelianas.

Otros estudios encaminados a determinar las consecuencias reales de las mutaciones en el ADN genómico sobre el ARN, como el ARN-Seq y otros estudios de transcriptómica y proteómica, permiten conocer el impacto transcripcional y bioquímico de las mutaciones y, por tanto, acercarse a la patogenia y fisiopatología de las enfermedades genéticas.

No siempre es sencilla la decisión de qué aproximación por NGS (paneles, WES, WGS) es la idónea. Dos factores principales determinan la elección de la técnica de NGS que elegir: la tasa de detección potencial de la técnica para el fenotipo interrogado, y la frecuencia de hallazgos inciertos o inesperados. La elección para el diagnóstico clínico de una de las técnicas de NGS debe estar determinada por la mayor capacidad de detección, el menor coste y la menor información incidental e incierta posibles. Esto se hace aún más importante en diagnóstico prenatal, que puede llevar a opciones reproductivas como la terminación de la gestación, con implicaciones éticas y legales inmediatas. Lo mismo ocurre con el diagnóstico preimplantacional. Los compromisos con el paciente del médico que le atiende o del investigador se establecen a través del asesoramiento genético y del proceso del consentimiento informado. El asesoramiento previo a la prueba genética es absolutamente necesario con estas técnicas. El clínico o el asesor genético adquiere el compromiso de gestionar los resultados del estudio, a través del asesoramiento posterior a la prueba genética y también mediante un programa de actuación médica concreto. El consentimiento informado debe también ajustarse a normas establecidas, y el paciente debe tener el derecho de revocación. La posibilidad de dar opciones a los pacientes para decidir si desean o no conocer ciertos tipos de información debe contemplarse. Un hecho importante es que las secuencias generadas por NGS, sobre todo por WES o WGS, siempre deben ser interpretadas a la luz del conocimiento en el momento del estudio.

De cara al futuro, es ciertamente previsible que la secuenciación vaya a estar presente en todos los ámbitos de la sanidad y en muchos otros de la vida de los ciudadanos. La rápida evolución de la tecnología de secuenciación hace previsible la generalización de nuevos métodos de secuenciación que incluirán la obtención de secuencias completas de grandes fragmentos de ADN –posiblemente cromosomas completos– de pequeños grupos de células o incluso células individuales. Los primeros prototipos operativos de este tipo de secuenciadores ya están disponibles. A su vez, es más que previsible que junto a la secuenciación de ADN se generalizará la secuenciación de ARN (ARN-seq), lo que permitirá monitorizar, de modo directo y continuo, estados fisiológicos y celulares. De cara al futuro es también relevante resaltar la importancia que muy pronto adquirirán las diferencias entre poblaciones humanas en el proceso de análisis de los datos genómicos.

Finalmente, mientras que el énfasis actual, particularmente en el ámbito más traslacional, se enfoca básicamente a la secuenciación de ADN y ARN, otros tipos de información son igualmente necesarios para poder extrapolar entre el nivel molecular y el fenotípico, enfermedad

y estados fisiológicos. Estos incluyen, entre otros: datos epigenéticos, tanto secuenciación de modificaciones de las bases del ADN (methyl-seq), como la secuenciación de elementos de control génico (chip-seq) como sitios de unión de factores de transcripción, regiones de unión de factores de modificación epigenético u otros factores que pueden modificar el nivel de expresión génica; datos de proteómica y metabolómica, cuyo obtención es todavía demasiado lenta y costosa, pero que sin duda serán necesarios en el contexto general del análisis de las bases moleculares de las enfermedades y además son potencialmente capaces de proveernos de mejores marcadores de enfermedades que los resultantes de la secuenciación de ADN.



Muestras y genomas habituales susceptibles de análisis: ADN germinal o constitutivo, ADN somático (tumoral y no tumoral) y metagenoma

En principio, cualquier muestra biológica que contenga ácidos nucleicos (ADN o ARN) es susceptible de ser analizada mediante técnicas de secuenciación masiva. Así, para el estudio y/o diagnóstico de la mayor parte de las enfermedades hereditarias (miocardiopatías hereditarias, cáncer hereditario, etc.) se utiliza como material de estudio ADN germinal obtenido de una muestra de sangre, aunque puede obtenerse a partir de otro tipo de muestras, como saliva, un raspado de mucosa, células (en cultivo, criopreservadas o inmortalizadas), bloques de tejido, etc.

En el caso del diagnóstico y/o estudio de enfermedades oncológicas, la muestra a secuenciar suele ser de origen tumoral (tumor fresco, criopreservado o fijado), células tumorales aisladas en el caso de tumores hematológicos, e incluso una muestra de sangre en el caso de tratar de identificar enfermedad mínima residual, que puede permitir la detección de una célula tumoral en 100 000 células. En la mayoría de los casos en los que se analiza una muestra tumoral se suele analizar también el ADN germinal con el objetivo de diferenciar qué cambios genómicos son somáticos (se han producido en el tumor) de aquellos cambios constitutivos (que están presentes en la línea germinal)¹⁷.

Por otra parte, el abaratamiento de los costes de secuenciación está haciendo posible el desarrollo de técnicas que permiten caracterizar las poblaciones de microorganismos (bacterias, virus, levaduras, hongos) presentes en determinadas localizaciones (cavidad bucal, aparato digestivo, aparato respiratorio, piel, genitales,...). Este tipo de estudios se conocen bajo el nombre de metagenómica, ya que se basan en el estudio del genoma de numerosos organismos diferentes (microbioma)¹⁸. Aunque los estudios metagenómicos son de gran interés en el campo de la microbiología y la industria alimentaria, el hecho de que un ser humano tenga más células bacterianas que células humanas, y la importancia de los microorganismos en el desarrollo de numerosas enfermedades humanas, hace de esta técnica un tipo de prueba con un alto potencial de crecimiento en los próximos años. Algunas de las ventajas de estas técnicas con respecto a los métodos microbiológicos empleados actualmente en clínica consisten en que tras una amplificación mediante PCR y secuenciación se puede obtener una respuesta en cuestión de horas, en comparación con la compleja realización de cultivos celulares, que pueden requerir varios días de complejos y costosos cultivos, y la posibilidad de detectar la presencia de microorganismos que no se pueden cultivar en los medios comúnmente utilizados. Aunque el tipo de genoma analizado en este tipo de estudios corresponde a genomas no humanos,

En cualquiera de los casos presentados anteriormente, la muestra biológica a secuenciar (ADN o ARN convertido a cADN) tiene que ser procesada para poder ser sometida al proceso de secuenciación. El proceso básico de preparación de una muestra suele consistir en tres pasos: i) fragmentación del genoma en fragmentos pequeños; ii) unión de unos adaptadores de ADN a dichos fragmentos, y iii) unión de estos fragmentos a un soporte sólido para su secuenciación. Dependiendo del tipo de estudio que se vaya a realizar, se pueden distinguir varios escenarios (tabla 2.1).

Tabla 2.1. Diversos niveles de análisis de estudios de NGS. Estos estudios pueden aplicarse tanto a genomas constitutivos como somáticos

Análisis	Explicación
Genoma	Se refiere al estudio de las variantes genéticas de la totalidad del genoma de un individuo. En un estudio clásico en genoma de ADN constitutivo humano se obtiene un número de variantes de aproximadamente 1-2 megabases comparado con el genoma de referencia. El método habitual de obtención es por fragmentación directa del ADN.
Exoma	Se refiere al estudio de las variantes genéticas de la totalidad de genes de un individuo. En un estudio clásico en exoma en ADN constitutivo humano (unos 23 000 genes; unos 180 000 exones), se obtiene un número de variantes de aproximadamente 50 000-80 000 bases comparado con el genoma de referencia. El método habitual de obtención es por captura y generación de librerías del ADN.
Exoma clínico	Se refiere al estudio de las variantes genéticas de los genes con implicación clínica actual ("genes OMIM") de un individuo. En un estudio clásico en exoma clínico en ADN constitutivo humano (unos 5 000 genes), se obtiene un número de variantes de aproximadamente 30 000-40 000 bases comparado con el genoma de referencia. El método habitual de obtención es por captura y generación de librerías del ADN.
Paneles	Se refiere al estudio de las variantes genéticas de los genes relacionados con una patología o un grupo de patologías en un individuo. Existen paneles de genes pequeños (de unos pocos genes, no más de 100) y paneles grandes de genes (de varios cientos o miles). El número de variantes comparado con el genoma de referencia depende del tamaño del panel. El método habitual de obtención es por captura y generación de librerías del ADN.
Estudio de amplicones	Se refiere al estudio de las variantes genéticas de los genes relacionados con una patología o un grupo pequeño de patologías, un gen o unos pocos genes en un individuo. Habitualmente de 1 a 5 genes. El número de variantes a analizar es menos de 100 en general. El método habitual de obtención es por amplificación mediante PCR y posterior generación de librerías del ADN.
Estudio de transcriptoma	Se refiere al estudio del conjunto de genes que se están expresando en un momento dado en una célula. La expresión de un gen supone que este ha sido transcrito a ARN mensajero. Las células de un mismo organismo y con un mismo genoma pueden llegar a ser tipos celulares muy dispares dependiendo de la combinación de genes que exprese cada una, lo que es lo mismo, dependiendo de su transcriptoma. El método habitual es la retrotranscripción de ARN a ADN y posterior captura y generación de librerías del ADN.

NGS: *Next Generation Sequencing.*

SECUENCIACIÓN DE GENOMA COMPLETO (WGS) O DE ARN (ARNseq)

El ADN de partida (o cADN para el ARNseq) se fragmenta en trozos de una longitud conocida (generalmente de 200 a 500 pares de bases). Luego se reparan estos fragmentos de ADN y se le unen los adaptadores de ADN específicos, y el ADN está listo para secuenciar. Este proceso se conoce como construcción de una librería genómica.

SECUENCIACIÓN DE EXOMA (WES) O DE PANEL DE GENES

La secuenciación del exoma completo o una selección de un panel de genes se realiza mediante enriquecimiento o captura. En muchas ocasiones no se necesita estudiar el genoma completo de un individuo o tumor, sino una serie de regiones de interés, que puede ser desde un gen o conjunto de genes (conocido como un panel de genes), o incluso todas las regiones codificantes de genes del genoma (conocido como exoma). Debido a que estas regiones representan alrededor del 1-2% del genoma, es necesario separar estas regiones de interés del resto del genoma, para proceder a su secuenciación. El procedimiento consiste en generar una librería genómica (v. párrafo anterior), pero a continuación se procede a seleccionar o capturar mediante una técnica de hibridación de ácidos nucleicos, aquellas regiones del genoma que se pretenden analizar. Una vez enriquecida la muestra en las regiones de interés, se procede a su secuenciación.

El llamado exoma clínico se refiere a un panel de genes muy grande que comprende las regiones codificantes de cerca de 5000 genes, de los que se conoce que producen patología mendeliana o monogénica conocida.

SECUENCIACIÓN DE GENES O PANELES MEDIANTE AMPLICONES

A diferencia de los anteriores, en este caso no se fragmenta el ADN, sino que se procede a la amplificación mediante PCR de las regiones de interés mediante la utilización de oligonucleótidos específicos que pueden contener o no los adaptadores necesarios para la secuenciación. Una vez obtenidos los fragmentos de ADN (conocidos como amplicones), y unidos los adaptadores, se pueden secuenciar como en los casos anteriores.

TÉCNICAS DE INMUNOPRECIPITACIÓN DE CROMATINA (ChIP-Seq), SECUENCIACIÓN BISULFITO Y SIMILARES (ChIA-PET, 3C, ETC.)

El desarrollo de las técnicas de secuenciación masiva ha permitido emplear estas técnicas para conocer cómo está organizado el genoma y saber qué proteínas interaccionan con qué regiones del genoma, conocer las modificaciones químicas que presenta, estudios de la organización de la cromatina y la disposición tridimensional del genoma en el núcleo. Aunque en todos estos estudios el objetivo está centrado en conocer cómo está organizado el genoma, y son irrelevantes las variaciones genéticas del individuo del que se obtiene la muestra para la

interpretación de los resultados, para obtener esta información es necesario secuenciar parte del genoma del individuo, por lo que los datos crudos generados permiten identificar variantes génicas del donante. Los procedimientos de preparación de las muestras para este tipo de estudios son muy variados y complejos y su descripción se aleja del objetivo de este documento. En cualquier caso, en los pasos finales lo que se obtiene es una serie de fragmentos de ADN que es necesario secuenciar para conocer a qué regiones del genoma pertenecen.

METAGENÓMICA

Como se ha mencionado anteriormente, estos estudios están centrados en conocer la diversidad de microorganismos que hay en una muestra (siendo las muestras biológicas humanas las de más relevancia para el propósito de este documento). La manera de abordar el análisis del metagenoma puede variar dependiendo del objetivo del estudio. La aproximación más habitual consiste en amplificar mediante una PCR el gen que codifica el rRNA 16S, un gen presente en todos los organismos vivos menos en virus. La amplificación de este gen mediante el empleo de oligonucleótidos específicos, seguido por la secuenciación masiva, permite determinar de manera cuantitativa, la abundancia de las distintas especies de bacterias presentes en una muestra. Dado que los oligonucleótidos empleados amplifican este gen en bacterias, ninguna de las secuencias obtenidas corresponde a ADN humano. Sin embargo, otras técnicas menos habituales, y encaminadas a identificar nuevos organismos que podrían perderse mediante la amplificación del rRNA 16S, se basan en una secuenciación masiva de muestras enriquecidas en microorganismos, pero que pueden contener secuencias de ADN humano, por lo que esta distinción es relevante a la hora del tratamiento de los datos derivados de estudios de metagenómica.

RESUMEN DEL CAPÍTULO

Cualquier muestra biológica que contenga ácidos nucleicos (ADN o ARN) es susceptible de ser analizada mediante técnicas de NGS. Así, para el estudio y/o diagnóstico de la mayor parte de las enfermedades hereditarias se utiliza como material de estudio ADN germinal obtenido de una muestra de sangre, aunque puede obtenerse a partir de otro tipo de muestras, como saliva, un raspado de mucosa, células (en cultivo, criopreservadas o inmortalizadas), bloques de tejido, etc. En el caso del diagnóstico y/o estudio de enfermedades oncológicas, la muestra a secuenciar suele ser de origen tumoral (tumor fresco, criopreservado o fijado), células tumorales aisladas en el caso de tumores hematológicos, e incluso una muestra de sangre en el caso de tratar de identificar enfermedad mínima residual o ADN circulante tumoral, que puede permitir la detección de una célula tumoral en 100 000 células.

El abaratamiento de los costes de secuenciación está haciendo posible el desarrollo de técnicas que permiten caracterizar las poblaciones de microorganismos (bacterias, virus, levaduras, hongos) presentes en determinadas localizaciones (cavidad bucal, aparato digestivo, aparato respiratorio, piel, genitales,...). Este tipo de estudios se conocen con el nombre de metagenómica, ya que se basan en el estudio del genoma de numerosos organismos diferentes (microbioma). Aunque los estudios metagenómicos son de gran interés en el campo de la microbiología e industria alimentaria, el hecho de que un ser humano tenga más células bacterianas que células humanas, y la importancia de los microorganismos en el desarrollo de numerosas enfermedades humanas, hace de esta técnica un tipo de prueba con un alto potencial de crecimiento en los próximos años.

En cualquiera de los casos presentados anteriormente, la muestra biológica a secuenciar (ADN o ARN convertido a cADN) tiene que ser procesada para poder ser sometida al proceso de secuenciación. El proceso básico de preparación de una muestra suele consistir en tres pasos:

- i) fragmentación del genoma en fragmentos pequeños;
- ii) unión de unos adaptadores de ADN a dichos fragmentos, y
- iii) unión de estos fragmentos a un soporte sólido para su secuenciación. Dependiendo del tipo de estudio que se vaya a realizar, se pueden distinguir varios escenarios:
 - a) secuenciación de genoma completo (WGS) o de ARN (RNAseq). El ADN de partida (o cADN para el ARNseq) se fragmenta en trozos de una longitud conocida (generalmente de 200 a 500 pares de bases). Luego se reparan estos fragmentos de ADN y se le unen los adaptadores de ADN específicos, y el ADN está listo para secuenciar. Este proceso se conoce como construcción de una librería genómica;
 - b) secuenciación de exoma (WES) o de panel de genes mediante enriquecimiento o captura. En muchas ocasiones no se necesita estudiar el genoma completo de un individuo o tumor, sino una serie de regiones interés, que puede ser desde un gen o conjunto de genes (conocido como un panel de genes), o incluso todas las regiones codificantes de genes del genoma (conocido como exoma);
 - c) secuenciación de genes o paneles mediante amplicones; a diferencia de los anteriores, en este caso no se fragmenta el ADN, sino que se procede a la amplificación mediante PCR de las regiones de interés mediante la utilización de oligonucleótidos específicos que pueden contener o no los adaptadores necesarios para la secuenciación;

- d) técnicas de inmunoprecipitación de cromatina (ChIP-Seq), secuenciación bisulfito y similares (ChIA-PET, 3C, etc.). El desarrollo de las técnicas de NGS ha permitido emplear estas técnicas para conocer cómo está organizado el genoma y saber qué proteínas interaccionan con qué regiones del genoma, conocer las modificaciones químicas que presenta, estudios de la organización de la cromatina y la disposición tridimensional del genoma en el núcleo, y
- e) estudios de metagenómica. Como se ha mencionado anteriormente, estos estudios están centrados en conocer la diversidad de microorganismos que hay en una muestra.



LOS DIFERENTES DATOS QUE SE OBTIENEN Y ARCHIVAN

NIVELES DE INFORMACIÓN. SIGNIFICADO Y RELEVANCIA CLÍNICA O PARA LA INVESTIGACIÓN BÁSICA. IDENTIFICABILIDAD DE LOS PARTICIPANTES

Datos no filtrados ni analizados

Una vez introducidas las librerías genómicas en los instrumentos de secuenciación, estos procederán a la secuenciación del ADN, generando de unos pocos miles a unos cientos o miles de millones de secuencias o lecturas. Una de las diferencias entre las distintas tecnologías de NGS (*Next Generation Sequencing*) es la longitud de las lecturas obtenidas, que puede variar varios órdenes de magnitud, desde unas pocas decenas de bases (50-100 bases) hasta más recientemente miles de bases cada lectura. Durante el proceso de secuenciación, el secuenciador suele generar unos datos crudos que en algunas de las tecnologías más empleadas se corresponden con simples imágenes de microscopía. Estos datos crudos tienen que ser procesados, generalmente por el mismo equipo de secuenciación, y al final se generan los datos de secuencias crudas y sin procesar.

Formato FASTQ. Aunque existen distintos formatos para almacenar estas secuencias crudas, el formato más estándar y más empleado es el denominado formato FASTQ (fig. 3.1). En este formato, cada lectura está representada por cuatro líneas:

1. La primera línea comienza por un símbolo "@" y contiene un identificador único de la lectura. Dado que un experimento de secuenciación puede generar mil millones de lecturas, y se almacenan todas en el mismo fichero FASTQ, es necesario asignar un identificador único a cada una de las lecturas.
2. La segunda línea contiene la secuencia de bases. Este campo es realmente el resultado de la secuenciación, con el orden de nucleótidos secuenciados.
3. La tercera línea comienza por un "+" y suele estar vacía o tener el identificador de la lectura.
4. La cuarta y última línea corresponde a la información de calidad de base. Esta línea proporciona información sobre el grado de fiabilidad de la secuencia de bases que se ha obtenido y que se muestra en la línea 2. Para cada base secuenciada y representada en la línea 2, hay

variantes estructurales (inserciones o deleciones de grandes fragmentos genómicos, translocaciones, inversiones,...). Desde el punto de vista práctico, cuando un investigador comparte el genoma de una persona/tumor con otro investigador, el fichero que se suele intercambiar es el BAM correspondiente al paciente en cuestión. Así mismo, cuando se deposita el genoma (o exoma o ARNseq) de una persona en un repositorio con acceso controlado (p. ej., EGA o dbGAP), el fichero que se deposita es el BAM.

Aunque el alineamiento al genoma de referencia es el método más extendido para el análisis de muestras de NGS, existen algunos procedimientos en los que es posible analizar un exoma o genoma sin necesidad de alinearlo a un genoma de referencia, y por lo tanto no requieren la generación de un fichero BAM. Estos procedimientos pueden consistir en una comparación directa del fichero FASTQ de una muestra frente a otra (p. ej., secuencias de un tumor frente a secuencias del tejido normal del mismo paciente), o bien en la generación de un ensamblaje *de novo*. Este último procedimiento es más complejo y aún no se ha utilizado para estudiar un genoma humano utilizando secuencias de NGS debido al pequeño tamaño de las lecturas generadas. Sin embargo, a medida que avancen las tecnologías de secuenciación y se consigan lecturas más largas, será factible el ensamblaje de genomas *de novo*, que no tienen por qué almacenarse en ficheros BAM, aunque podrían utilizarse formatos similares para comparar con el genoma de referencia.

Datos filtrados y analizados. Niveles de análisis y tipo de variabilidad

El formato BAM almacena todas las lecturas alineadas contra el genoma de referencia, por lo que contiene la información necesaria para llevar a cabo los distintos tipos de análisis que se pueden realizar. Dependiendo del objetivo del estudio, el tipo de análisis de estos datos, así como el tipo de filtros a utilizar, será distinto. No obstante, el tipo de análisis más frecuente, encaminado a la identificación de variantes génicas causantes de una determinada patología, suele requerir al menos tres niveles de análisis:

1. **Extracción de variantes genómicas.** El primer nivel consiste en identificar todas las variantes en las que el genoma de la muestra secuenciada difiere de la secuencia del genoma de referencia. El número de variantes obtenido al analizar un genoma completo es del orden de 4 millones, y el de un exoma completo de aproximadamente 50 000.
2. **Filtrado de variantes.** El segundo nivel consiste en filtrar las variantes para tratar de eliminar aquellas que no proporcionan información relacionada con la enfermedad estudiada. El tipo de filtro a utilizar varía mucho dependiendo del objetivo del análisis.
3. **Identificación de variantes funcionales.** Una vez seleccionadas todas las variantes de los pasos anteriores, el último nivel de análisis de datos consiste en definir cuál o cuáles de todas esas variantes tiene un efecto funcional. El objetivo en este caso puede ser diagnosticar una enfermedad debido a la presencia de una mutación en un gen responsable de esta enfermedad, o en el caso de un genoma tumoral, identificar mutaciones que afecten al pronóstico o a la respuesta a un tratamiento, y por tanto, tengan una relevancia clínica.

Por lo tanto, una vez secuenciado un panel, un exoma o un genoma y alineado al genoma de referencia en formato BAM, el siguiente paso consiste en identificar las variantes genéticas y almacenarlas en un tipo de formato estándar. El formato más empleado para almacenar la información de variantes genéticas es el formato VCF (*Variant Calling Format*) (fig. 3.3)¹⁹.

Este formato permite describir de manera sencilla la mayor parte de las variantes génicas, proporcionando información acerca del locus genómico en el que se localiza la variante (cromosoma y posición), qué secuencia es la presente en el genoma de referencia y cuál es la que se ha observado en el genoma analizado, y luego otros parámetros como el genotipo, y una puntuación que refleja la confianza que se tiene en la variante. A continuación se describe la información relevante para las variaciones de un solo nucleótido.

```
9 139390648 . CAG C 152 . INDEL;IS=10,0.476190;DP=21;VDB=1.268478e-01;AF1=0.5;AC1=1;DP4=3,2,5,4;MQ=51;FQ=15
5;PV4=1,7.5e-05,0.02,1 GT:PL:DP:GQ 0/1:190,0,222:14:99

9 139391636 . G A 222 . DP=63;VDB=1.945541e-01;RPB=-7.586541e-01;AF1=0.5;AC1=1;DP4=9,15,8,14;MQ=51;FQ=223;
PV4=1,0.43,1,1 GT:PL:DP:GQ 0/1:252,0,255:46:99
```

Figura 3.3. Formato VCF. Ejemplo de dos variantes genéticas en el cromosoma 9, identificadas por NGS y representadas en formato VCF.

Las variaciones de un solo nucleótido (SNV, *single nucleotide variant*), o sustituciones, son el tipo de variante más frecuente en el genoma, con entre 3 y 4 millones de sustituciones en cada persona cuando se compara con el genoma de referencia. La mayor parte de estas sustituciones son polimórficas en la población (con una frecuencia alélica superior al 1%), por lo que se conocen como SNP (*single nucleotide polymorphisms*). Sin embargo, no todas las variantes de un solo nucleótido son polimorfismos, ya que la mayor parte de las enfermedades hereditarias o mutaciones en cáncer se deben a este tipo de variante genética.

A partir de los datos presentes en un fichero BAM es posible identificar los SNV presentes en el genoma de un individuo, y es posible determinar si se encuentran en homocigosis (los dos alelos tienen el mismo cambio) o en heterocigosis (un alelo tiene el cambio y el otro no). Sin embargo, la identificación de SNV a partir de un fichero BAM no es un proceso trivial, sino que la capacidad de identificar un SNV depende de varios parámetros de la secuenciación, así como del contexto de secuencia genómica en el que se encuentra. Debido a la trascendencia que tiene este punto para la fiabilidad de los diagnósticos basados en estas tecnologías, se describirá brevemente el parámetro que más influye en la capacidad de detectar SNV.

El parámetro más crítico para poder identificar cualquier variante génica es la cobertura. Este parámetro refleja cuántas lecturas se han obtenido correspondientes a la región de interés. Es decir, si al secuenciar el genoma de un individuo hay una región (p. ej., el exón 2 de *BRCA1*) que no se ha secuenciado bien, es posible que no haya ni una sola lectura correspondiente al exón 2. Por lo tanto, si no hay lecturas es imposible concluir si hay variantes o si no las hay en esa región. Si por el contrario hay una única lectura, y esa lectura tiene un cambio, es posible que ese cambio sea un SNV en el exón 2 de *BRCA1*. Sin embargo, y como se ha comentado anteriormente, todas estas tecnologías cometen errores de secuenciación, por lo que es imposible afirmar de manera concluyente que haya un cambio cuando solo hay una lectura de cobertura. Si por el contrario hay varias lecturas (3 o 4) y todas tienen el mismo cambio, se tendrá más seguridad acerca del cambio en esa posición del genoma, y a medida que se aumente la

cobertura, más seguro se estará. De igual forma, aunque con cuatro lecturas se pueda tener cierta confianza de que hay un cambio en el genoma, ya que las cuatro lecturas muestran dicho cambio, esa cobertura es insuficiente para conocer el genotipo, es decir, si los dos alelos (el materno y el paterno) tienen el cambio (se encuentra en homocigosis) o solo un alelo (heterocigosis). La situación es análoga a tirar una moneda al aire: la probabilidad de que una vez tirada la moneda, en las siguientes tres tiradas salga lo mismo es de un 12,5%, que es una probabilidad muy alta. Por lo tanto, el que salga cara en cuatro tiradas consecutivas no quiere decir que la moneda no tenga cruz. En conclusión, a mayor cobertura, mayor capacidad para detectar SNV y determinar el genotipo con suficiente confianza. Sin embargo, el aumento de la cobertura supone un mayor gasto en secuenciación, por lo que siempre se intenta encontrar un balance entre ambos parámetros (cobertura y coste). Por ello, es importante determinar si el estudio de NGS que se va a realizar está dentro del contexto de la investigación o de la asistencia/diagnóstico, y en genomas constitutivos o somáticos, ya que dependiendo de esto las coberturas necesarias son diferentes y las variables de análisis distintas.

Una consecuencia de la complejidad de identificar variantes dependiendo de factores como la cobertura es que la identificación de una variante genética mediante este tipo de técnicas de NGS se realiza con un cierto grado de incertidumbre. Por esta razón, para cada variante identificada es necesario establecer algún parámetro probabilístico que refleje la confianza estadística que se tiene en dicha afirmación. El formato VCF permite describir una variante incluyendo toda esta información adicional, así como el tipo de filtros que se ha utilizado posteriormente, por lo que es el formato en el que se suelen representar todos los datos de variantes procesadas.

Sin embargo, el número de variantes que pueden identificarse al secuenciar un exoma o un genoma es muy elevado, y la mayoría de estas variantes no tienen un efecto causal en la patogénesis de la enfermedad estudiada. Por esta razón, una vez identificadas todas las variantes, es necesario llevar a cabo un filtrado de los datos para reducir el número total de variantes, eliminando las no informativas, y almacenando las informativas. Este sería el segundo nivel de organización de los datos. El tipo de filtros aplicados a los datos va a variar dependiendo del tipo de análisis del que se trate. Algunos de los más habituales se mencionan a continuación:

- **Genomas del cáncer.** La secuenciación de un genoma tumoral tiene por objetivo identificar aquellos cambios presentes en el genoma de las células tumorales y no en las células no tumorales del mismo paciente. Para ello es necesario comparar las variantes del genoma tumoral con las variantes del genoma normal del mismo paciente, y eliminar todas aquellas variantes que ya estaban en la línea germinal del individuo. Mediante la utilización de algoritmos específicos y una serie de filtros es posible definir un conjunto de mutaciones somáticas específicas del tumor, que suele oscilar desde unos pocos cientos de variantes hasta cientos de miles.
- **Enfermedades hereditarias de causa conocida.** En el caso de querer identificar una enfermedad hereditaria de la que se conoce el gen o genes responsables de la enfermedad, el análisis se centra solo en las variantes localizadas en las regiones que contienen los genes de interés, y dependiendo del número de genes, el número de variantes puede oscilar entre unas pocas y unas miles. Dichas variantes se filtran mediante la utilización de bases de datos que contienen mutaciones causantes de dicha enfermedad, con el objetivo de tratar de identificar alguna mutación patogénica previamente descrita, y en el caso de no identificarse mutacio-

nes conocidas, suelen aplicarse filtros que permitan eliminar polimorfismos frecuentes en la población, para tratar de definir un número reducido de variantes para estudios posteriores.

- **Enfermedades hereditarias de causa desconocida.** Este grupo de patologías es el que más se ha beneficiado del desarrollo de las nuevas tecnologías de secuenciación. El análisis es relativamente complejo, ya que el número de variantes es muy elevado, y la causa genética, normalmente única, desconocida. Frecuentemente, en estos casos se tiene información acerca del tipo de herencia (dominante, recesiva, ligada al sexo, mitocondrial), lo que condiciona el tipo de filtro que se va a emplear. Los dos primeros filtros empleados suelen ser la eliminación de polimorfismos genéticos frecuentes en la población, ya que si la patología es poco frecuente, no puede estar causada por una variante genética frecuente en la población; y en segundo lugar, se suele limitar el estudio a las variantes que afectan a genes, ya que constituye la parte del genoma mejor conocida y en la que se han identificado más del 90% de las mutaciones causantes de enfermedades humanas. No obstante, el número de candidatos que pasan estos filtros suele situarse en torno a 100-200, lo que es un número excesivamente elevado. Por esta razón, el segundo tipo de filtro que suele aplicarse es secuenciar muestras de los padres y/o familiares adicionales, con el objeto de descartar variantes que no sigan el patrón de herencia o que estén presentes en familiares que no presentan dicha patología. Al final de este proceso de filtrado se habrán generado unos resultados que consisten en una serie de variantes (generalmente menos de 30), que podrían ser las causantes de la enfermedad. Luego continuará el proceso de determinar si estas variantes están presentes en poblaciones controles y poblaciones de pacientes con la patología que se está estudiando y si segregan con las enfermedades dentro de la familia.

IDENTIFICABILIDAD DE LOS FICHEROS

En cualquier caso, los ficheros generados, tanto con los datos crudos como con los filtrados, no contienen información clínica acerca del paciente secuenciado. No obstante, los datos crudos y las variantes sin filtrar contienen toda la información genómica de una persona, lo que podría permitir su identificación si los datos genéticos de dicha persona, o de algún familiar cercano, están depositados en otra base de datos. En concreto, un individuo puede participar en un estudio en el que se hayan analizado un número alto de marcadores genéticos (bien por secuenciación de genoma, exoma, ARNseq, genotipado o panel de varios genes). Esta información se deposita en una base de datos. Si dicho individuo participa en otro estudio independiente mediante secuenciación de genoma o exoma, un investigador podría comparar los datos genómicos de dicho individuo con la otra base de datos y determinar, con un alto grado de significación estadística, que los dos conjuntos de datos pertenecen al mismo individuo. Pero en ningún caso permite que el investigador, a partir de estos datos, sea capaz de determinar la identidad del donante.

La única información adicional que podría recabar un investigador al comparar los datos de dichos estudios sería algún tipo de información clínica o demográfica de la que quizás no dispusiera. Un ejemplo puede ser que la secuenciación del genoma que se esté haciendo se relacione con una enfermedad A, y los datos presentes en la otra base de datos se refieran a otra enfermedad B. Esta comparación permitiría conocer una información (que el individuo en cuestión sufre o sufrió la enfermedad B) de la que no disponía el investigador. Por otra parte, el hallazgo de una variable genética muy muy infrecuente podría permitir la identificación de un individuo específico. Un caso distinto es la identificación de familiares cercanos a partir de datos genómicos, y la posible imputación de variantes patogénicas en su genoma sin haber realizado una prueba genética específi-

ca para ello. Un ejemplo sencillo sería el de una persona a la que se le realiza un estudio de NGS y se le identifica una mutación en un gen de susceptibilidad a cáncer hereditario. El investigador puede tener acceso a datos de genotipado de otro estudio, por ejemplo un estudio de asociación génica (GWAS) de enfermedades neurológicas, de test de memoria, etc. Al comparar los datos del genoma del primer paciente con la base de datos de este estudio es posible detectar que un pariente del primer paciente ha participado en el segundo estudio (puede ser descendiente o ascendiente directo, o bien primos de segundo, tercer o cuarto grado). La participación de una persona en un estudio GWAS puede ser como caso (persona con enfermedad) o como control (persona sana), aunque en este caso es irrelevante.

Un dato a tener en cuenta es que en este tipo de estudios el genotipado se realiza mediante el análisis de arrays de SNP, que determinan polimorfismos comunes en la población. A partir de este tipo de análisis es imposible detectar si una persona tiene una mutación patogénica en un gen de cáncer hereditario, ya que solo se estudian polimorfismos comunes en la población. Sin embargo, la comparación del genoma del primer paciente con esta base de datos ha revelado que un familiar (p. ej., una prima) ha participado en dicho estudio GWAS. Debido a la manera en la que se hereda la información genética, es posible determinar qué regiones cromosómicas tienen en común ambos familiares, y puede que una de las regiones compartidas corresponda a la región que contiene el gen de cáncer hereditario. Las variantes genéticas que se encuentran cerca en un cromosoma suelen heredarse juntas, y constituyen lo que se llama haplotipos. A partir del análisis de estos dos tipos de datos (secuenciación de genoma y SNP de genotipado), es posible determinar qué haplotipos comparten los dos familiares. Si la secuenciación del genoma completo ha revelado la existencia de una mutación en un gen de cáncer hereditario en el primer paciente, y se puede establecer en qué haplotipo se encuentra dicha mutación, y dicho haplotipo es el que comparte con su prima, sería posible imputar, con un alto grado de confianza, que la prima tiene o no tiene la misma mutación, a pesar de que no se le ha hecho una prueba genética para dicho gen, y aunque a partir de los datos crudos de genotipado no se pueda extraer este tipo de información. Estos resultados indican que esta persona puede tener dicha mutación, con una probabilidad elevada. No obstante, dado que no se ha secuenciado dicho gen en esta persona, esta sospecha de mutación ha de ser confirmada mediante los procedimientos habituales. Todas estas posibilidades deben tenerse en cuenta a la hora de determinar si los datos que se van a obtener deben ser considerados como datos de carácter personal desde la perspectiva jurídica. En el capítulo 4 (v. pág. 53) se explica que la razonabilidad en el esfuerzo para la identificación del sujeto es un criterio determinante a estos efectos.

Las implicaciones éticas y jurídicas de la gestión de los hallazgos inesperados (información que no se buscaba y que no se sabía que se iba a obtener) son complejas, como se verá más adelante. Sin embargo, en este caso no se ha identificado tal mutación mediante el análisis de los datos genómicos del sujeto, pero se tiene una sospecha muy alta de que la persona pueda tener dicha mutación. La única manera de confirmar la presencia de esta es realizando una prueba específica. Por el contrario, para la persona implicada también constituye una decisión muy trascendente. Así, dicha persona ha participado en un estudio (quizás como control sano), en el que se analizaban polimorfismos frecuentes en la población, y en ningún momento se estaba realizando una prueba que pudiera revelar datos referentes a la salud de dicha persona. Sin embargo, los datos de secuenciación del genoma de un familiar han revelado la posible presencia de una mutación patogénica en su genoma, pero su confirmación exige la realización de una prueba diagnóstica, ya que los datos de imputación no son suficientes para confirmar el diagnóstico molecular.

Aunque el hecho de utilizar información genética para el diagnóstico suele generar un nivel de cautela elevado, la situación anterior se podría asemejar a situaciones más habituales en clínica. Un análisis sanguíneo rutinario para controlar el nivel de colesterol revela la presencia de un número elevado de linfocitos, lo que podría indicar la presencia de un tumor hematológico, pero es necesario llevar a cabo pruebas adicionales. Dado que se trata de un hallazgo inesperado, y que se necesitan pruebas adicionales para la confirmación diagnóstica, la situación sería similar a la descrita anteriormente, con la salvedad de que la prueba utilizada se realizó para participar en un estudio. De ahí la importancia de conocer la opinión del paciente o donante ante posibles hallazgos que puedan tener implicaciones patogénicas para el propio individuo y su familia. En las páginas 78 y 90 se explica cómo se resuelve esta situación en la legislación española, y en la página 64 se describe el papel que pueden desempeñar los comités de ética asistencial.

VARIANTES CON INFORMACIÓN DE UTILIDAD CLÍNICA

En realidad, y a pesar de los progresos realizados en biología molecular y genética/genómica, el número de variantes de utilidad clínica o accionables, entendida la utilidad clínica como que el conocimiento de poseer una mutación concreta puede conducir a pautas de tratamiento específicas o al empleo de fármacos selectivos, es todavía muy pequeño. En el caso de cáncer puede decirse que en el mejor de los casos se han caracterizado unas decenas de mutaciones de las que se pueden deducir posibles pautas clínicas. Así mismo disponemos solo de unas decenas de fármacos aprobados para su utilización y sujetos a una regulación sobre casos de uso muy estricta.

Es, por tanto, prioritario incrementar la cantidad de información genética disponible para su utilización en términos médicos. Estos incluyen, por una parte, la mejora de la tecnología de validación de resultados genómicos, tanto a nivel puramente técnico, incluyendo la confirmación de las mutaciones detectadas, como por su consistencia a nivel de las enfermedades; esto es, por la presencia de las mutaciones en un número significativo de pacientes. También es importante mejorar la metodología usada para determinar sistemáticamente la significación de las mutaciones en términos bioquímicos y fisiológicos, puesto que estos estudios funcionales requieren un conocimiento específico de cada caso e implican un conjunto complejo de aproximaciones experimentales. El uso de modelos animales, por ejemplo el ratón o el pez cebra, constituyen la alternativa más extendida para realizar un primer estudio de las consecuencias fenotípicas de las mutaciones, mientras que el uso de modelos animales más complejos, como los modelos de ratones portadores de xenoinjertos de tumores (ratones avatar), son una alternativa cada vez más utilizada en fases más avanzadas de los estudios fenotípicos y de respuesta a fármacos, sobre todo en estudios de investigación preclínicos.

Finalmente, el incremento del número de mutaciones y variantes genéticas relevantes clínicamente dependerá de la capacidad de llevar a cabo ensayos clínicos para evaluar la efectividad de distintos tratamientos a pacientes con las mismas alteraciones genéticas, así como de los métodos bioinformáticos para transformar la información disponible en predicciones fiables. De esta manera se podrá evaluar tanto las consecuencias de las mutaciones como las mejores opciones de tratamiento disponibles para cada caso.

Toda esta problemática es particularmente relevante en el caso de la interpretación de los hallazgos inesperados que inevitablemente aparecen en cualquier estudio genómico²⁰. El estudio de los hallazgos inesperados presenta un conjunto de dificultades técnicas, prácticas y éticas,

especialmente en cuanto a cómo y en qué grado debe informarse al afectado. Una de las dificultades principales es que habitualmente no se informa de la posibilidad de tales hallazgos, ya que no es el objetivo del estudio, y el problema surge cuando estos se producen (como se ha señalado antes, en las páginas 78 y 90 se explica cómo se resuelve esta situación en la legislación española, y en la página 64 se describe el papel que pueden desempeñar los comités de ética asistencial), los comités de ética de la investigación o los comités de ética externos de los biobancos). Diferentes estudios han demostrado que las intervenciones que promueven expectativas realistas de la secuenciación del genoma pueden atenuar las respuestas negativas a la incertidumbre²¹.

De hecho, cada vez conocemos más casos de mutaciones consideradas de modo general como patogénicas que, sin embargo, se encuentran presentes en individuos sanos que las toleran quizás por tener un fondo genético distinto con un conjunto de mutaciones compensatorias o un sistema de regulación génico/fisiológico diferentes. Por tanto, de un modo general puede decirse que la predicción de las consecuencias de las mutaciones encontradas accidentalmente requiere, además de la aplicación de los métodos generales para interpretación de las mutaciones aisladas, el análisis integral del contexto genómico y fisiológico específico de cada caso.

No todas las variantes potencialmente identificadas tienen la misma implicación y relevancia para la salud de un individuo. Así, en un estudio genómico podemos encontrar: mutaciones que confirman la enfermedad que provocó el análisis del paciente, mutaciones que son clínicamente relevantes y cuyo conocimiento supone un beneficio directo para el paciente o su descendencia (ya sea porque se puede intervenir sobre ellas o bien porque, aunque no tienen tratamiento, su conocimiento puede ayudar al paciente a tomar decisiones reproductivas), mutaciones de alta penetrancia relacionadas con enfermedades de aparición tardía (con y sin tratamiento en la actualidad), mutaciones que no causan enfermedad en el paciente pero que pueden causarla a su descendencia (portadores de enfermedades autosómicas recesivas o portadoras de enfermedades ligadas al cromosoma X), mutaciones que confieren una cierta predisposición a padecer una determinada enfermedad compleja, o variantes y cambios relacionados con caracteres de base genética de dudosa relación con la salud humana. La utilidad y necesidad de realizar pruebas para la detección de estas variantes ha sido hasta ahora dependiente de la opinión personal de los profesionales dedicados a la genética y es, por lo tanto, muy heterogénea.

Informes clínicos

No existe una pauta unificada acerca de la información que debe incluirse en los informes que resultan de un análisis genómico. Tampoco en cuanto a los hallazgos inesperados que pueden llegar a producirse con este tipo de tecnologías. En el American College of Medical Genetics (ACMG) se acordó por consenso, en el año 2014, incluir dentro del informe 56 mutaciones con repercusión clínica asociadas a 24 enfermedades en las que existe evidencia de que su detección puede prevenir o mejorar resultados médicos adversos graves, pero esta iniciativa no ha sido adoptada por muchos centros europeos y ha generado una enorme polémica¹¹.

Los informes de un laboratorio clínico deben seguir unas normas para garantizar la calidad del resultado, para que el profesional solicitante y el paciente puedan entender el resultado

y su fiabilidad, y para permitir la transmisión de información a través del consejo genético y dirigir la atención médica especializada. Las nuevas técnicas de secuenciación dan lugar a una enorme cantidad de variantes de secuencia y la información y evidencia científica sobre sus implicaciones es aún incompleta y lo seguirá siendo durante años, por lo que la incertidumbre puede ser una consecuencia importante de su uso. Por otro lado, los hallazgos incidentales ya mencionados pueden tener implicaciones clínicas relevantes, a veces mucho más importantes que la respuesta al motivo de solicitud de la secuenciación. El laboratorio debe por tanto ser capaz de reflejar en su informe todos aquellos datos resultantes de la secuenciación con potenciales implicaciones clínicas y también aquellos con consecuencias aún inciertas para los que no podemos demostrar su carácter de variación normal. Para ello, es habitual realizar un análisis y consulta en bases de datos poblacionales, de enfermedades específicas y de secuencias de genomas consenso que son útiles en el análisis de datos de estudios de NGS, tanto asistenciales como de investigación (tabla 3.1).

Deben constar en los informes, y hacerlo de forma comprensible para el receptor, principalmente el profesional solicitante y también el paciente, los siguientes datos:

- El motivo clínico de la solicitud que permita dar respuesta al solicitante a una pregunta concreta. ¿Tiene el paciente la mutación A que confirmaría su enfermedad y permitiría utilizarla para el consejo genético?
- El tipo de estudio: diagnóstico, predictivo, de portador, prenatal, de informatividad, farmacogenético, etc.
- La metodología empleada, su sensibilidad y especificidad, la cobertura, y el resto de datos que permitan conocer la calidad (en términos de validez analítica, validez clínica) y la fiabilidad de la información.
- En los resultados:
 - Todos los datos relevantes del análisis de la secuencia estratificando su relevancia clínica. No es aceptable que el profesional receptor no sepa identificar las características de cada uno de los hallazgos:
 - Hallazgo de mutación para la enfermedad sospechada clínicamente.
 - Hallazgo de mutaciones patogénicas causantes de otras enfermedades con carácter de presentación inmediata, predictivo y/o presintomático.
 - Hallazgo de variantes inciertas de las que no se conoce su carácter de variación normal y que por su posición genómica y la predicción de la función del gen concreto o por niveles de evidencia bajos (estudios animales, celulares) que permiten sospechar patogenicidad, deban ser notificadas.
 - La interpretación de estos resultados en términos de relevancia clínica:
 - La mutación identificada que confirma la enfermedad.
 - Otras mutaciones que pueden causar una enfermedad con mayor o menor probabilidad, abordando los conceptos de penetrancia, susceptibilidad, variabilidad en la expresión e incertidumbre.
 - Un cambio incierto puede hacer sospechar un riesgo de salud potencial y es tributario de seguimiento clínico.
 - La implicación de los resultados en términos de manejo médico y de consejo genético. Existen dos modelos posibles:
 - La referencia a la necesidad de seguimiento médico y consejo genético sin más explicaciones.

Tabla 3.1. Algunas bases de datos poblacionales, de enfermedades específicas y de secuencias de genomas consenso que son útiles en el análisis de datos de estudios de NGS, tanto asistenciales como de investigación

Bases de datos poblacionales	Web	Comentarios
Exome Aggregation Consortium	http://exac.broadinstitute.org/	Base de datos de las variantes encontradas durante la secuenciación del exoma de 61 486 individuos no relacionados secuenciados como parte de diversos estudios genéticos de enfermedades específicas y de poblaciones. Se excluyeron los sujetos con enfermedad pediátrica, así como los individuos emparentados.
Exome Variant Server	http://evs.gs.washington.edu/EVS	Base de datos de las variantes encontradas durante la secuenciación de exomas de varias cohortes grandes de individuos de ascendencia americana, europea y africana. Incluye los datos de cobertura.
1000 Genomes Project	http://browser.1000genomes.org	Base de datos de las variantes (a baja cobertura) encontradas durante la secuenciación genómica específica de 26 poblaciones. Proporciona una mayor diversidad en comparación con el servidor Exome Variant, pero también contiene datos de menor calidad, y algunas cohortes contienen individuos relacionados.
dbSNP	http://www.ncbi.nlm.nih.gov/snp	Base de datos de las variaciones genéticas cortas (típicamente ≤50 pb) obtenida y enviada de muchas fuentes diferentes. Pueden carecer de los detalles del estudio de origen y pueden contener variantes patogénicas.
dbVar	http://www.ncbi.nlm.nih.gov/dbvar	Base de datos de la variación estructural (normalmente > 50 bp) obtenida y enviada de muchas fuentes diferentes.
Bases de datos de enfermedades específicas	Web	Comentarios
ClinVar	http://www.ncbi.nlm.nih.gov/clinvar	Base de datos de hallazgos sobre la importancia y el fenotipo clínico de la relación de variantes humanas.
OMIM	http://www.omim.org	Base de datos de genes y enfermedades humanas. También contiene una muestra representativa de las variantes genéticas asociadas con las enfermedades.
Human Gene Mutation Database	http://www.hgmd.org	Base de datos de las variantes y anotaciones publicadas en la literatura. Requiere suscripción de pago para tener acceso a la mayor parte del contenido, aunque hay una versión gratuita con menos información.
Human Genome Variation Society	http://www.hgvs.org/dblist/dblist.html	Sitio de la Human Genome Variation Society en la que se incluyen bases de datos con anotaciones de variantes en subgrupos específicos. Un gran porcentaje de las bases de datos se construyen en el sistema abierto de base de datos de Leiden.
Leiden Open Variation Database	http://www.lovd.nl	
DECIPHER	http://decipher.sanger.ac.uk	Base de datos de citogenética molecular para los clínicos y los investigadores que incluye los datos de microarrays genómicos con fenotipos utilizando el navegador Ensemblgenome.
Bases de datos de secuencias	Web	Comentarios
NCBI Genome	http://www.ncbi.nlm.nih.gov/genome	Repositorio de secuencias de genoma de referencia.
RefSeqGene	http://www.ncbi.nlm.nih.gov/refseq/rsg	Herramienta para secuencias de genes de referencia.
Locus Reference Genomic (LRG)	http://www.lrg-sequence.org	
EGA	https://www.ebi.ac.uk/ega/home www.ncbi.nlm.nih.gov/gap	Donde se depositan las secuencias de individuos secuenciados en estudios.
MitoMap	http://www.mitomap.org/MITOMAP/ HumanMitoSeq	Base de datos de secuencias de referencia de ADN mitocondrial humano de la Universidad de Cambridge.

- o La explicación que incluya elementos detallados de las implicaciones del resultado, con referencia a las fuentes que han generado la evidencia:
 - Para el pronóstico y tratamiento del paciente.
 - En cuanto a la forma de herencia y los riesgos asociados a la mutación.
 - Para un posible diagnóstico prenatal o preimplantacional en términos de opciones reproductivas disponibles.
 - Para otros miembros de la familia con riesgo de tener la misma variante y sus implicaciones clínicas.

Interpretación de los resultados

Como ya se ha comentado, los estudios genómicos identificarán un enorme número de variantes, y muchas de ellas con implicaciones en la salud de un individuo. A esto hay que sumar que previsiblemente muchas de las variantes que hoy por hoy son de significado incierto o desconocido pueden tener en un futuro próximo una significación patológica. Con los conocimientos médico-genéticos actuales se estima que la media de variantes relacionadas con enfermedad que podrían afectar a un individuo se sitúa en torno a 300, y algunos autores vaticinan que en el futuro las variantes genómicas con implicaciones más o menos relevantes en patología humana pueden llegar a 10 000²².

Por esta razón, entre otras, como puede ser “la falsa sensación de salud” si a alguien se le comunica que no tiene una determinada variante de predisposición a una enfermedad compleja o la generación de una ansiedad innecesaria si se comunica la presencia de esa misma variante, es conveniente hacer un filtrado de la información que se obtiene en los estudios genómicos informando tan solo de aquellos que tengan relevancia y sean de utilidad clínica para el paciente (incluyendo como de utilidad clínica aquellas mutaciones sobre las que no se puede intervenir médicamente pero que el paciente presenta en estado de portador y pueden causar enfermedad en su descendencia) y siempre y cuando el paciente dé su consentimiento para ser informado de ellas tratando de no vulnerar su autonomía y el derecho a “no saber”.

Muy recientemente se han publicado artículos consenso sobre la interpretación, priorización y clasificación de una variante hallada en un estudio de secuenciación^{23,24}.

En este consenso, se clasifica a las variantes en patogénicas y benignas, se establecen los cinco niveles clásicos de las variantes (patogénica, probablemente patogénica, benigna, probablemente benigna y variante de significado incierto) y se asignan características específicas de priorización de cada variante para poder encuadrarlas en alguno de los cinco niveles (tabla 3.2). Es probable que esta clasificación se generalice y se establezca como estándar de uso habitual en la clasificación de variantes. Con los análisis genómicos amplios, estas variantes ya no tendrán que ser buscadas de forma activa, sino que aparecerán por defecto. En este sentido, la reciente Orden SSI/2065/2014 de cartera de servicios comunes del Sistema Nacional de Salud y el procedimiento para su actualización, incluyendo las pruebas genéticas (31 de octubre de 2014) contiene la cartera mínima de pruebas genéticas y pretende marcar las directrices que deben seguirse en cuanto a la realización de las pruebas genéticas en el SNS. Específicamente pretende “concretar los aspectos fundamentales de la cartera de genética definiendo el consejo genético y los grupos de patologías para los que se indicará dicho consejo genético, así como los tipos de análisis genéticos y los criterios para la indicación de cada uno de ellos”.

Tabla 3.2. Clasificación habitual de las variantes genómicas según su patogenicidad*

Criterios para la clasificación de variantes patológicas	
Evidencia de patogenicidad	Categoría
Muy grave	Variantes "nulas" PVS1: (nonsense, frameshift, regiones canónicas +/- 1-2 sitios de <i>splicing</i>, codones de iniciación, deleciones de uno o varios exones) en un gen donde se conoce que existe LOF (pérdida de función) como mecanismo patogénico. Advertencias: ■ Tener cuidado con los genes donde la pérdida de función no es un mecanismo conocido de la enfermedad (p. ej., GFAP, MYH7). ■ Tener cuidado al interpretar variantes con pérdida de función en el extremo 3' de un gen. ■ Tener cuidado con las variantes de <i>splicing</i> que se prevé que conduzcan a la deleción de un exón, pero deja el resto de la proteína intacta. ■ Tener precaución en presencia de múltiples transcritos.
Grave	PS1: cambio de aminoácido igual que una variante patógena previamente establecida, independientemente de los cambios de nucleótidos. Ejemplo: Val Leu → causada por cualquiera de G > C o G > T en el mismo codón. Advertencia: Tener cuidado con los cambios que afectan al <i>splicing</i> más que en el nivel de aminoácido/proteína. PS2 <i>de novo</i> en un paciente con la enfermedad y que no tiene antecedentes familiares (cuando la paternidad y maternidad hayan sido confirmadas). Nota: la confirmación de la paternidad sola es insuficiente. La donación de óvulos, la maternidad de alquiler, los errores en la transferencia de embriones, y otros ejemplos, pueden contribuir a la no maternidad. PS3 bien establecida mediante estudios funcionales de apoyo (<i>in vitro</i> o <i>in vivo</i>) de un efecto perjudicial en el producto/s del gen/genes. Nota: los estudios funcionales que han sido validados y que han demostrado ser reproducibles y robustos se consideran los mejores. PS4 La prevalencia de la variante en los individuos afectados es significativamente mayor en comparación con la prevalencia en los controles. Nota 1: el RR o el OR (obtenido a partir de estudios de casos y controles) es > 5,0, y el intervalo de confianza del RR o del OR no incluye 1. Nota 2: en los casos de variantes muy raras en los que los estudios de casos y controles no alcanzaron significación estadística, la observación previa de la variante en muchos pacientes no relacionados con el mismo fenotipo, y su ausencia en controles, se pueden utilizar como nivel moderado de certeza.
Moderada	PM1 Situado en un punto caliente (<i>hot spot</i>) mutacional y/o dominio funcional crítico y bien establecido (p. ej., el sitio activo de una enzima) sin variación benigna. PM2 Ausente en controles (o una frecuencia extremadamente baja si es recesivo) en Exome Variant Server, 1000 Genomes o EXAC. Advertencia: Los datos de población para inserciones/deleciones pueden estar no suficientemente representados en los estudios de NGS. PM3 Para los trastornos recesivos, detectados en <i>trans</i> con una variante patógena. Nota: Esto requiere pruebas de los padres (o descendientes) para determinar la fase (<i>cis</i> o <i>trans</i>). PM4 Cambios en la longitud de proteínas como resultado de deleciones/inserciones en el marco de lectura en una región sin repeticiones o variantes "stop-loss". PM5 Nuevo cambio missense en un aminoácido, donde otro cambio missense ha demostrado ser patógeno en estudios anteriores. Ejemplo: Arg156His es patógeno; ahora se observa Arg156Cys. Advertencia: Tener cuidado con los cambios que afectan al <i>splicing</i> más que aminoácidos/proteínas. PM6 Asumido como <i>de novo</i>, pero sin confirmación de la paternidad y maternidad.
Adicionales	PP1 Cosegregación de la variante con la enfermedad en varios miembros de la familia afectados en un gen conocido que definitivamente causa la enfermedad. Nota: Se puede utilizar como una evidencia más fuerte a medida que aumenta la segregación de datos y el número de familiares. PP2 Missense en un gen que tiene una baja tasa de variación missense benigna y en el que las variantes missense (y no las nonsense) son un mecanismo común de la enfermedad. PP3 Múltiples líneas de evidencias bioinformáticas apoyan un efecto perjudicial sobre el gen o el producto génico (conservación, impacto del <i>splicing</i>, etc.). Advertencia: debido a que muchos de los algoritmos <i>in silico</i> utilizan el mismo (o muy similar) algoritmo de entrada para sus predicciones, cada algoritmo no debe ser considerado como un criterio independiente. PP3 se puede utilizar una sola vez en toda evaluación de una variante. PP4 El fenotipo del paciente o la historia familiar es altamente específico para una enfermedad con una única etiología genética. PP5 Una publicación u otra fuente de confianza informa recientemente que la variante es patológica, pero la evidencia no está disponible para el laboratorio para llevar a cabo una evaluación independiente.

LOF (*loss of function*): pérdida de función; PM: *pathogenic moderate* (variante moderada). PP: *pathogenic additional* (variante patológica adicional); PS: *pathogenic strong* (variante severa); PVS: *pathogenic very strong* (variante muy severa).

*Habitualmente se utilizaba un sistema de cinco niveles que reconoce variantes como benigna, probablemente benigna, de significado incierto, probablemente patológica, y patológica. Sin embargo, este sistema de puntuación es, en su mayor parte, aplicable solo a trastornos mendelianos. Si se consideran todos los factores genéticos y no genéticos, tales como en las enfermedades humanas complejas, también existen otras categorías para definir la relación entre una variante genética y la enfermedad (variante de susceptibilidad, factor pronóstico, y modificador de la respuesta a fármacos, etc., entre otros).

Tabla 3.2. Clasificación habitual de las variantes genómicas según su patogenicidad* (cont.)

Criterios para la clasificación de variantes benignas	
Evidencia de benignidad	Categoría
Definitiva	BA1 La frecuencia alélica es > 5% en Exome Sequencing Project, 1000 Genomes o EXAC (Exome Aggregation Consortium).
Fuerte	BS1 La frecuencia del alelo es mayor de lo esperado para la enfermedad. BS2 Observado en individuos adultos sanos para un trastorno recesivo (homocigotos), dominante (heterocigoto), o trastorno ligado al cromosoma X (hemocigoto) con penetrancia completa esperada a una edad temprana. BS3 Estudios funcionales bien establecidos <i>in vitro</i> o <i>in vivo</i> no muestran ningún efecto perjudicial sobre la función de proteínas o del <i>splicing</i>. BS4 Falta de segregación en los miembros afectados de una familia. Advertencia: La presencia de fenocopias para fenotipos comunes (es decir, cáncer, epilepsia) puede imitar la falta de segregación entre los individuos afectados. Además, las familias pueden tener más de una variante patógena que contribuya a un trastorno autosómico dominante, además de la confusión de una aparente falta de segregación.
Adicional	BP1 Variante missense en un gen para el que se conocen variantes que principalmente truncan la proteína como causa de la enfermedad. BP2 Observado en <i>trans</i> con una variante patógena de un gen dominante/trastorno totalmente penetrante u observado en <i>cis</i> con una variante patógena en cualquier patrón de herencia. BP3 Deleciones o inserciones que rompen el marco de lectura en una región repetitiva sin una función conocida. BP4 Múltiples líneas de evidencia bioinformáticas sugieren que la variante no tiene ningún impacto en el gen o producto génico (conservación, <i>splicing</i>, etc.). Advertencia: debido a que muchos de los algoritmos <i>in silico</i> utilizan el mismo (o muy similar) algoritmo de entrada para sus predicciones, cada algoritmo no debe ser considerado como un criterio independiente. BP4 se puede utilizar una sola vez en toda evaluación de una variante. BP5 Variante encontrada en un caso con una base molecular alternativa para la enfermedad. BP6 Una publicación fiable o una fuente de confianza informa recientemente de que se trata de una variante benigna, pero la evidencia no está disponible para el laboratorio para llevar a cabo una evaluación independiente. BP7 Una variante sinónima (silente) para la que los algoritmos de predicción de <i>splicing</i> no predicen ningún impacto ni la creación de un nuevo sitio de <i>splicing</i> y el nucleótido no está altamente conservado.

BA: Benign Stand-Alone (variante definitivamente benigna); BP: Benign Supporting (variante benigna adicional); BS: Benign Strong (variante fuertemente benigna).

Clasificación de la variante	Explicación
Patogénica	(i) 1 Muy grave (PVS1) y (a) ≥ 1 Grave (PS1-PS4) o (b) ≥ 2 Moderadas (PM1-PM6) o (c) 1 Moderada (PM1-PM6) y 1 adicional (PP1-PP5) o (d) ≥ 2 Adicionales (PP1-PP5) (ii) ≥ 2 Graves (PS1-PS4) o (iii) 1 Graves (PS1-PS4) y (a) ≥ 3 Moderadas (PM1-PM6) o (b) 2 Moderadas (PM1-PM6) y ≥ 2 adicionales (PP1-PP5) o (c) 1 Moderada (PM1-PM6) y ≥ 4 adicionales (PP1-PP5)
Probablemente patogénica	(i) 1 Muy grave (PVS1) y 1 moderada (PM1-PM6) o (ii) 1 Grave (PS1-PS4) y 1-2 moderadas (PM1-PM6) o (iii) 1 Grave (PS1-PS4) y ≥ 2 adicionales (PP1-PP5) o (iv) ≥ 3 Moderadas (PM1-PM6) o (v) 2 Moderadas (PM1-PM6) y ≥ 2 adicionales (PP1-PP5) o (vi) 1 Moderada (PM1-PM6) y ≥ 4 adicionales (PP1-PP5)
Benigna	(i) 1 Variante poblacional (BA1) o (ii) ≥ 2 Graves (BS1-BS4)
Probablemente benigna	(i) 1 Graves (BS1-BS4) and 1 adicional (BP1-BP7) o (ii) ≥ 2 Adicionales (BP1-BP7)
Variante de significado incierto	(i) Otros criterios no mencionados arriba o (ii) Los criterios para benignidad y patogenicidad son contradictorios

Interacción con los profesionales solicitantes

La interacción entre el médico solicitante y los laboratorios que realizan la prueba es muy variable dependiendo de qué tipo de centro y laboratorio se trate, y dependiendo de si se trata de estudios asistenciales o de investigación. Así, en un extremo estarían aquellos equipos clínicos especializados en genética o con otros especialistas que tienen formación suficiente en genética y que forman parte del mismo centro/servicio/laboratorio en el que se realizan las pruebas, por lo que no parece necesaria la inclusión en el informe más que la indicación de la necesidad de consejo genético. Si, en el otro extremo, es un profesional no genetista quien solicita el estudio a un laboratorio distante, la necesidad de información clara y detallada se hace fundamental. Por ello, debe tenerse en cuenta que los especialistas tratantes de los pacientes con algunas enfermedades de base genética proporcionan consejo médico y necesitarán, si no tienen el entrenamiento y capacitación pertinente, de la ayuda de otros profesionales que proporcionen consejo genético. Esto hace que los profesionales del laboratorio deban tener una formación suficiente en consejo genético y genética clínica y que, en ocasiones, deberán incluir en plantilla asesores genéticos capaces de mejorar la información aportada en los informes.

Por último, al igual que con el consejo genético, la recomendación de seguimiento médico por uno o varios especialistas como consecuencia de un hallazgo genético puede estar explicitada en el informe del laboratorio. Una vez más requiere que el servicio o unidad de genética disponga de personas con el suficiente nivel de formación clínica para ello. Todo esto parece contemplado en el mencionado Real Decreto 639/2014, donde se prevé la formación de genetistas clínicos (clínica y laboratorio) en España, recientemente aprobado dentro de la ley de Troncalidad, aún pendiente de desarrollo.

Los informes clínicos elaborados por los especialistas en genética clínica, en asesoramiento genético y/o en otras consultas de especialistas con suficiente formación en genética deben detallar las implicaciones para el paciente concreto y su familia, de forma personalizada, y deben acompañarse de las sesiones de información necesarias para comunicar de forma eficiente las implicaciones de la información genética/genómica. Cuanto mayor respaldo y confianza reciba del laboratorio que ha realizado y emitido un informe de NGS, más segura y eficaz será su actuación.

ARCHIVO Y GESTIÓN DE LA INFORMACIÓN. NIVELES DE ACCESO

El manejo de datos derivados de secuenciación masiva de muestras humanas puede tener unos requerimientos muy distintos dependiendo de que su uso sea asistencial, o por el contrario, se vayan a usar en investigación. Así, una prueba diagnóstica se basa en un procedimiento conocido y aprobado para dicho fin, y por lo general solo se realiza una vez por muestra, dando lugar a un informe clínico. Por el contrario, en el caso de la investigación es frecuente tratar de desarrollar nuevos procedimientos o probar determinadas hipótesis de trabajo. Esto suele conllevar que una misma muestra se analice numerosas veces cambiando determinados parámetros, lo que supone un consumo de poder de procesamiento mucho mayor que en el ámbito asistencial. Además, en determinados ámbitos de investigación es muy frecuente la colaboración, lo que permite a grupos analizar las muestras procedentes de distintos centros de investigación o reanalizar muestras utilizadas en otros estudios. El uso de dichos datos genómicos, siempre aprobados por los correspondientes comités de ética, hace que el número

de paneles/exomas/genomas a los que puede tener acceso un grupo de investigación sea muy superior a los que se pueden generar en un hospital a lo largo de un año, lo que requiere una capacidad de almacenamiento mucho mayor que en el caso asistencial.

Debido a estas características, en este documento se han tratado de manera separada las necesidades para el análisis y almacenamiento en el ámbito asistencial, de aquellas en el ámbito de la investigación biomédica. De hecho, aunque en ambos casos hay necesidades de procesamiento y almacenamiento, así como de acceso a los datos, la naturaleza de la investigación hace que las necesidades, y por lo tanto, los costes asociados a la investigación, puedan ser muy superiores a los requeridos en un ámbito asistencial. En cualquier caso, los modelos organizativos son similares en ambos casos. En este apartado se discuten los principales modelos de archivo y gestión de datos genómicos en el ámbito clínico, finalizando con un punto referente al ámbito de la investigación.

Por otra parte, no se puede perder de vista que, con frecuencia, existen conexiones estrechas entre la asistencia y la investigación que se reflejan en las necesidades relativas en la obtención, almacenamiento y utilización de los datos. Sería interesante que los biobancos desempeñaran un papel en este sentido, pero esto no parece que vaya a suceder a corto plazo.

El Real Decreto 1716/2011, de 18 de noviembre, por el que se establecen los requisitos básicos de autorización y funcionamiento de los biobancos con fines de investigación biomédica y del tratamiento de las muestras biológicas de origen humano, define un “biobanco con fines de investigación biomédica” como aquel “establecimiento público o privado, sin ánimo de lucro, que acoge una o varias colecciones de muestras biológicas de origen humano con fines de investigación biomédica, organizadas como una unidad técnica con criterios de calidad, orden y destino, con independencia de que albergue muestras con otras finalidades”. El mismo real decreto plantea la posibilidad de la existencia de una “red de biobancos”, plataforma promovida por el Instituto de Salud Carlos III y que agrupa a 52 instituciones distribuidas en 14 comunidades autónomas²⁵.

Existen en la actualidad diversas situaciones en los biobancos en cuanto al tratamiento y almacenamiento de los datos asociados a las muestras; sin embargo, de forma generalizada, los biobancos españoles no almacenan los datos genómicos de las muestras que gestionan. Generalmente, cuando un grupo de investigación solicita unas muestras a un biobanco para un estudio genómico, los datos genómicos no vuelven al biobanco, ni siquiera los datos crudos. Es el grupo de investigación el que financia el estudio genómico y el que gestiona posteriormente estos datos. Sería interesante que los biobancos pudiesen disponer de los datos genómicos asociados a las muestras; sin embargo, tal y como está organizado el sistema actualmente resulta inimaginable, por un lado, que un biobanco encargue la generación de los datos genómicos de sus muestras (por costoso), y por otro lado, que los grupos de investigación que generan estos datos genómicos los pusiesen a disposición de los biobancos una vez acabado el estudio, sin ningún tipo de contraprestación, aunque la financiación pública de la secuenciación debería ser un factor que favoreciera esta posibilidad.

Archivo y niveles de acceso a la información en asistencia

ARCHIVO EN HISTORIA CLÍNICA

Desde el punto de vista clínico, los análisis genómicos constituyen una prueba diagnóstica que ayuda en la toma de decisiones médicas. Al igual que la mayoría de pruebas realizadas, los

datos de una prueba genética deben incorporarse a la historia clínica del paciente. No obstante, el volumen de datos generado por los estudios genómicos hace que el archivo y gestión de estos datos sea distinto al de otras pruebas diagnósticas, por lo que es necesario valorar diferentes modelos de gestión de estos.

El archivo en la historia clínica del paciente de los datos crudos de la secuenciación genómica parece poco factible, ya que la necesidad de memoria y velocidad de procesamiento parece muy grande para los servidores encargados hoy de gestionar las historias clínicas de un centro o de un sistema de salud en nuestro país. Así, a pesar la creciente implantación de la historia clínica digital, los datos genómicos en los que se basa dicho informe (ficheros de secuencias en formato BAM) son difícilmente incorporables a dicha historia debido al elevado volumen de datos (hasta 100 gigabytes de información para un genoma completo). Sin embargo, los informes de laboratorio, que pueden contener de forma más o menos completa la lista de variantes de secuencia de un individuo y la interpretación de estas, así como los informes clínicos y de consejo genético, sí que podrían incorporarse a la historia clínica. En la mayoría de sistemas de archivo de historias clínicas implantado se incluye toda la información disponible del paciente, sin barreras, ya que se considera, por un lado, que puede ser relevante para sus cuidados médicos, y por otro, que los profesionales con acceso harán un uso responsable de este.

Existen, sin embargo, características de la información genética que la hacen particularmente tributaria de confidencialidad y de respeto singular:

- Los datos identifican estados de portador o enfermo “para toda la vida”, no problemas transitorios que serán simplemente un antecedente médico una vez superados.
- Los datos identifican enfermos, pero también portadores sanos, estados de salud presintomáticos o predicciones de riesgos de enfermedad.
- Los datos se generan para el paciente, pero pueden tener implicaciones para el resto de familiares.
- La complejidad de la información a transmitir con frecuencia no se corresponde con la formación en genética del profesional solicitante de la prueba. Por tanto, la gestión de los sistemas de información de los servicios de salud deberá adaptarse para garantizar el cumplimiento de la legislación vigente y la de futuro desarrollo, y deberá abordar una serie de aspectos importantes que actualmente no están resueltos, y hacerlo de forma común para los distintos servicios de salud de un país y probablemente de la Unión Europea. Algunos de estos aspectos son:
 - Quién puede solicitar la prueba.
 - Quién puede informar el resultado del estudio (v. cap. 6).
 - Si este resultado puede ser utilizado para la información de un familiar y con qué proceso de autorización, o con qué indicaciones médicas concretas (v. pág. 59).
- Si un dato que identifica estados presintomático, predictivo o de portador debe constar en la historia clínica de un paciente, dado que se trata de información que puede ser relevante en un futuro. Pero esto deberá valorarse en tanto en cuanto la posibilidad de establecer acciones preventivas o terapéuticas y la finalidad y pertinencia del estudio.

La Ley 41/2002, de Autonomía del Paciente, representa un marco general que puede dar respuesta a algunos de estos puntos, puesto que prevé la obligación general de que toda

la documentación clínica se archive en la historia y de que los profesionales accedan a los datos cuando sea pertinente por razones asistenciales. No obstante, es cierto que cuantitativa, y en cierto sentido también cualitativamente, la información genética representa retos en relación con su almacenamiento y condiciones de acceso. Podría ser oportuno por estas razones establecer criterios específicos a los que se presta atención en epígrafes posteriores.

Dejar en manos de los profesionales el control de la información de libre acceso para ellos parece ingenuo; les supone unos conocimientos sólidos de genética clínica y un conocimiento detallado de la ley y sus implicaciones. Deben ser las instituciones y servicios de salud quienes desarrollen los mecanismos para garantizar los derechos de usuarios y profesionales respecto a los datos genómicos.

ALMACENAMIENTO DE DATOS GENÓMICOS CRUDOS

En cuanto a la necesidad de almacenamiento de los datos genómicos crudos, existen dos grandes corrientes con respecto a cuál es la mejor manera de proceder. Así, la primera de ellas aboga por que, una vez emitido el informe clínico, se destruyan los datos genómicos, evitando así el problema de almacenamiento. Esta postura tiene a su favor el continuo abaratamiento de los costes de secuenciación, de tal manera que en algunas situaciones es posible que la repetición de una prueba de secuenciación sea una alternativa más económica que el almacenamiento de datos indefinidamente. No obstante, y dado el limitado conocimiento que se tiene hoy en día de la influencia de las variantes genéticas en numerosas patologías, el avance en este campo irá ligado a la disponibilidad de los datos genómicos de miles o millones de pacientes, por lo que su destrucción limitaría la capacidad de avanzar en estos aspectos. Un ejemplo lo constituye el estudio de muestras tumorales. En este contexto, puede ser necesario secuenciar muestras tumorales en distintas etapas del desarrollo de la enfermedad (biopsia inicial, biopsia tras recidiva, metástasis, etc.). En el caso de recidivas, la resistencia a un tratamiento puede deberse a la presencia de una mutación que ya estaba presente en una subpoblación de células tumorales en el tumor primario. Sin embargo, frecuentemente dichas subpoblaciones son muy minoritarias (inferiores al 5% del total de células del tumor). En estas condiciones, aunque la mutación esté presente en un subclón, la capacidad de detectar dicha alteración es tan baja, que en la mayor parte de las ocasiones no se incluirá en el informe clínico, ya que la evidencia estadística que sustenta la presencia de dicha alteración es demasiado baja. Sin embargo, una vez secuenciada la recidiva e identificada una alteración, se pueden reanalizar los datos del tumor primario para confirmar si dicha alteración genética estaba presente en el tumor primario. Así, aunque dicha alteración esté por debajo del límite de detección considerado para el informe clínico inicial, es posible detectarla en un segundo análisis, por lo que la destrucción de los datos originales impedirán realizar este tipo de análisis, que pueden proporcionar información relevante desde el punto de vista clínico referentes a respuestas a tratamientos.

En resumen, existen varias posibilidades de archivo de los ficheros que surgen de estudios de secuenciación masiva, que pueden resumirse en la tabla 3.3, y que se comentan en el punto siguiente.

Tabla 3.3. Diversos entornos y posibilidades de almacenamiento de los datos genómicos en distintos ámbitos (investigación/asistencia) y en diversas instituciones o centros

Sitios y modelos de almacenamiento de ficheros resultantes de estudios de NGS	
Local estricto	En servidores <i>ad hoc</i> localizados en el propio servicio/unidad del solicitante o del grupo que realiza los estudios de secuenciación.
Local compartido	En servidores localizados en una centralizada unidad de informática o similar, del centro del solicitante o del grupo que realiza los estudios de secuenciación.
Local extendido	Almacenamiento en un servidor central del organismo/agencia de investigación, en forma centralizada o de un servicio sanitario de toda una comunidad.
Externo	Almacenamiento en la nube (<i>cloud computing</i>) en servidores externos a las instituciones intervinientes, dentro o fuera del país.

ARCHIVOS EN SERVIDORES DE LA INSTITUCIÓN. MODELOS ORGANIZATIVOS

El archivo de datos genómicos procedentes de pruebas de diagnóstico asistencial requiere una planificación rigurosa debido al volumen de datos que se generan (hasta 100 gigabytes por genoma completo), a la demanda prevista para dichos análisis a medio y largo plazo, y a la necesidad de garantizar la privacidad de los datos. Por esta razón, y a pesar de que los análisis genómicos pueden ser solicitados por profesionales pertenecientes a distintos servicios, es aconsejable que el archivo de los datos se lleve a cabo de manera centralizada. Debido a la infraestructura y conocimientos necesarios para hacer frente a estas necesidades, parece adecuado que el servicio de informática de la institución sea la unidad responsable del archivo de datos genómicos. Esto permitirá crear los sistemas adecuados para facilitar el acceso controlado a los distintos profesionales implicados en el análisis de datos genómicos, así como que dichos sistemas sean escalables, para dar respuesta al previsible aumento en el volumen de este tipo de datos en la tarea asistencial.

Así mismo, debe contemplarse la posibilidad de que análisis realizados dentro de un marco asistencial puedan ser utilizados posteriormente en investigación si así lo ha autorizado el paciente (v. pág. 85), por lo que será necesario emplear herramientas y protocolos que faciliten el intercambio de datos. En este sentido, la experiencia de los biobancos en la gestión de muestras biológicas para su uso en investigación, así como su conocimiento de la legislación aplicable para cada tipo de solicitud, puede ser de gran utilidad para gestionar la cesión de datos genómicos con fines de investigación.

Como se ha mencionado anteriormente, el volumen de datos generados en cualquier estudio de secuenciación (genoma completo, exoma o paneles) es demasiado grande como para almacenarlo en ordenadores personales. Además, es necesario distinguir entre la capacidad de computación, necesaria para generar los datos procesados, y la capacidad de almacenamiento, requerida para almacenar los datos derivados de los distintos pacientes. Así, la capacidad computacional requerida para convertir las lecturas crudas (FASTQ) al formato BAM, y la identificación posterior de variantes (archivos VCF), no es un proceso trivial, por lo que es necesario disponer de servidores especializados, con gran

capacidad de computación, así como con suficiente capacidad de almacenamiento. Las necesidades de procesamiento dependerán de:

- a) **Tipo de análisis utilizado.** Los paneles de genes requieren poca capacidad de procesamiento, los exomas, media, y los genomas completos tienen unos requerimientos de procesamiento muy elevados. Dependiendo de cuál sea la distribución de estos tipos de análisis en un determinado centro hospitalario, las necesidades computacionales variarán sustancialmente.
- b) **Número de pacientes esperado para cada análisis y la periodicidad de llegada de los datos.** Obviamente, un parámetro que influirá en la elección de la capacidad computacional es el número de pacientes esperados para cada tipo de análisis utilizado, y si los datos crudos llegan periódicamente, o si casi todos llegan en bloque, lo que puede provocar la existencia de cuellos de botella en el análisis de los datos.
- c) **Tiempo de diagnóstico.** Por último, la capacidad de computación va a determinar el tiempo requerido para poder concluir el análisis genómico y emitir un informe clínico. Por esta razón, la elección de una capacidad de computación u otra dependerá del tipo de análisis utilizado y del tiempo en el que se desea obtener dicho resultado. Así, en el caso de pruebas genéticas de las que dependa la elección de un tratamiento específico para el paciente (como puede ser en el caso de un tumor), es deseable disponer del informe en el menor tiempo posible.

Teniendo en cuenta estos tres parámetros, y asumiendo la utilización de un mismo análisis bioinformático, será posible determinar los distintos escenarios y los requerimientos computacionales necesarios para conseguir los objetivos establecidos. En cualquier caso, y debido al previsible aumento en la demanda de este tipo de análisis, será necesaria la utilización de sistemas escalables que permitan adaptarse progresivamente a la demanda.

En cuanto a los modelos organizativos, se pueden distinguir dos grandes modelos para el manejo de datos de secuenciación para asistencia. Dichos modelos se diferencian atendiendo a si los datos se almacenan y procesan en servidores pertenecientes a la institución o si por el contrario se realiza en equipos de terceros (*cloud computing*).

Archivo en servidores locales

El uso de servidores locales es por ahora la opción más extendida, debido a que la novedad de las técnicas de NGS hace que el número de hospitales con acceso a datos de genomas completos o exomas sea todavía muy limitado. Además, en la mayor parte de los casos se suele tratar de grupos estrechamente ligados a la investigación básica, por lo que suelen contar con capacidad de análisis al disponer de infraestructura para computación intensiva derivada de sus proyectos de investigación. No obstante, desde el punto de vista organizativo, así como a lo largo de este documento, resulta recomendable la separación de los servidores dedicados a análisis de tipo asistencial, de aquellos destinados a la investigación básica, ya que los costes y las consideraciones de acceso restringido a distintos profesionales son muy dispares en ambas situaciones.

En relación con el archivo y procesamiento en servidores locales se debe tener en cuenta que:

- La decisión de archivo de los datos genéticos corresponde al grupo que realiza la investigación; actualmente son pocas las instituciones que están en disposición de ofrecer a sus gru-

pos espacios de archivo en servidores institucionales habilitados para este fin (a excepción de las que constituyen *per se* una infraestructura de supercomputación o secuenciación).

- Pese a varios intentos de dimensionar las necesidades de archivo de estos datos, es muy difícil realizar este análisis, incluso para una única institución, debido al incremento exponencial de la necesidad de terabytes (TB) de capacidad y a la deslocalización de estos.
- El coste de los servidores y de su almacenamiento puede ser considerable.
- Por último, no es sencillo distinguir entre servidores locales o en la nube ya que, en la práctica, pese a almacenar los datos en servidores locales, numerosas casas de NGS ponen a la disposición de los genetistas herramientas de análisis *user-friendly* que en multitud de ocasiones están la nube. Es decir, que en la gran mayoría de casos se tienen modelos mixtos (local-nube) en el binomio archivo-análisis de datos genómicos.

Aparte de la capacidad de procesamiento, que se suele utilizar para generar el fichero BAM y extraer las variantes genómicas presentes en dicho genoma, el archivo de los datos genómicos a largo plazo requiere sistemas de almacenamiento dedicados. Las dimensiones de estos sistemas para el archivo de la información se pueden calcular fácilmente y dependen fundamentalmente de los puntos a) y b) anteriores, es decir, del tipo de análisis utilizado (panel/exoma/genoma completo) y del número de pacientes analizados. Un ejemplo ilustrativo es la secuenciación del genoma completo de un tumor. Normalmente, este tipo de análisis requiere la secuenciación del genoma tumoral, así como del genoma normal del mismo paciente. El volumen de datos de este estudio particular se reparte de la siguiente manera:

- **Secuencias crudas (FASTQ): 0,2 TB.** La mitad correspondiente a las lecturas del tumor, y la otra mitad, a las de las células no tumorales.
- **Ficheros BAM: 0,2 TB.** La mitad para cada uno de los dos genomas secuenciados (tumor y normal).
- **Fichero VCF y otros ficheros de datos analizados y filtrados: 1 GB (gigabyte).**

En este tipo de estudio sería necesario almacenar los ficheros BAM (0,2 TB) y el fichero VCF (1 GB), mientras que los datos de mutaciones somáticas, de mucho menor tamaño, se podrían incorporar a la historia clínica del paciente. En el caso de los paneles de genes, la cantidad de información generada es mucho menor (menos de 1 GB), por lo que es posible almacenar y analizar los datos generados en equipos más sencillos. En la práctica, si un centro hospitalario realizase 500 genomas tumorales al año, necesitaría una capacidad de almacenamiento a largo plazo de 100 TB cada año, mientras que si lo que hiciera fueran 500 exomas, solo necesitaría 5 TB, y si solo realizase 500 paneles génicos, requeriría tan solo 0,5 TB. Por lo tanto, las necesidades son muy variables, pero dado que cada tipo de análisis tiene una indicación diagnóstica diferente, la situación más probable será la existencia de una combinación de los tres tipos de análisis, por lo menos a medio plazo. En cualquier caso, sería deseable la existencia de un único sistema de archivo de estos datos, ya que minimiza los costes de mantenimiento y facilita el empleo de herramientas comunes para el acceso a los datos o su cesión para investigación.

En cualquier caso, dentro de este apartado se pueden considerar varios modelos organizativos:

- a) **Análisis y almacenamiento en servidores propios.** Como se mencionaba anteriormente, un modelo básico consistiría en la utilización de servidores gestionados por el servicio de

informática de la institución, en el que se llevarían a cabo los análisis bioinformáticos de las distintas pruebas genómicas que se soliciten, y en el que también se almacenarían los datos genómicos. Como se explicará más adelante (v. pág. 83) el sistema debería contar con una persona autorizada que gestionara los accesos en función de las necesidades asistenciales. Por otra parte, el disponer de un servidor local permite instalar las herramientas de análisis bioinformático más adecuadas atendiendo a los criterios de sensibilidad y especificidad que consideren los profesionales implicados, así como modificar los parámetros de acuerdo a las necesidades del servicio. Además, en el caso de datos genómicos asistenciales que se vayan a utilizar también en investigación, el administrador podrá atender a una petición del biobanco solicitando que determinados datos genómicos se compartan con un investigador que cumple todos los requisitos.

No obstante, dicho modelo también tiene una serie de desventajas. Así, la institución tiene que desembolsar por adelantado la infraestructura informática, con la incertidumbre de cuál será su grado de utilización a medio plazo, dado que no existen precedentes del uso de este tipo de pruebas y la evolución de estas tecnologías todavía es muy rápida. Por otra parte, es necesaria la existencia de profesionales cualificados que sean capaces de decidir qué tipo de análisis se van a desarrollar, qué herramientas utilizar, y qué modelo de informe clínico es el más adecuado para no desbordar con información no útil clínicamente. Además, han de ser capaces de detectar posibles fallos en los protocolos de análisis bioinformáticos, y de establecer pautas de control de calidad para asegurar el correcto funcionamiento del servicio implementado en su centro. Y por último, se han de establecer los mecanismos que permitan el acceso controlado a los distintos niveles de información.

b) Análisis en servidores ajenos y almacenamiento en servidores propios. Una alternativa al punto anterior consiste en externalizar el análisis de los datos a un proveedor de servicios, de tal forma que una vez que se obtienen las secuencias, estas son transferidas a un servidor donde se procesan los datos, se genera un fichero BAM y un análisis de resultados, e incluso un informe clínico, que son devueltos al hospital. Los datos genómicos pasan a ser almacenados en los servidores institucionales, y a partir de los resultados se genera un informe clínico, si no se ha obtenido ya del proveedor de servicios.

La ventaja de este sistema consiste en que todo el procesamiento de datos se externaliza, por lo que no es necesario invertir en capacidad de procesamiento, solo en almacenamiento. Es posible que determinados proveedores de servicios se especialicen y tengan sistemas de análisis más sensibles o más específicos que los disponibles para implantar en un hospital, lo que constituye una ventaja clara, aunque este punto tiene que ser estudiado antes de decidirse por un proveedor determinado. Al igual que en el caso anterior, es necesario contar con profesionales cualificados que sean capaces de establecer controles de calidad para conocer si los servicios ofrecidos cumplen con los mínimos exigidos por el propio servicio, e informar de ello a los médicos que van a utilizar dichas pruebas diagnósticas.

Un aspecto fundamental para el empleo de este tipo de soluciones externalizadas por parte de un centro hospitalario es la existencia de una buena infraestructura de telecomunicaciones que permita la transferencia de datos a alta velocidad. Así, la mayor parte de los hospitales públicos cuentan con conexiones de acceso a Internet deficientes, que solo permiten la utilización de herramientas de bajo consumo de datos (correo electrónico, navegación de páginas web, descarga de ficheros de tamaño moderado,...). Sin embargo, la transferencia de los ficheros correspondientes a un genoma humano desde un centro hospitalario hasta un proveedor de servicios no es un proceso trivial. Así, dependiendo del tipo de conexión, estas

transferencias pueden llegar a durar varios días o incluso semanas, optándose en algunos casos por enviar mediante mensajería discos duros con dichas secuencias. Por esta razón, la ventaja que supone la externalización del servicio de análisis se puede perder si no se dispone de una infraestructura de telecomunicaciones adecuada, por lo que los responsables de valorar las distintas opciones deberán tener en cuenta si la capacidad de conexión del centro es suficiente para una opción externalizada.

Por otra parte, la utilización de un proveedor de servicios exige el establecimiento de unos términos de servicio y cláusulas de confidencialidad en cuanto a asegurar la privacidad de los datos, el tiempo que los datos estarán en servidores ajenos al hospital, y determinar las responsabilidades y acciones si la calidad de los análisis no se ajusta a lo determinado en los términos de servicio (v. pág. 59).

Archivo en la nube y modelos de almacenamiento remoto

Durante los últimos años se ha producido un abaratamiento de los precios de secuenciación, lo que ha facilitado el acceso a este tipo de estudios. Sin embargo, el precio de los equipos informáticos así como del almacenamiento de datos no ha descendido de manera proporcional, lo que está provocando que el análisis de datos se convierta en uno de los principales gastos de este tipo de estudios. Como se ha comentado anteriormente, el análisis de datos genómicos requiere el uso de equipos con gran capacidad de computación. Tanto su establecimiento como su mantenimiento y actualización requieren una considerable inversión económica. El aprovechamiento de dicha inversión es mayor cuanto mayor uso tengan dichos servidores. Por esta razón, en el caso de instituciones que no tengan una demanda muy alta de computación, o en el caso de centros que necesiten puntualmente una capacidad computacional muy superior a la disponible en su institución (p. ej., un hospital que suele realizar paneles de genes para diagnóstico, pero necesita analizar genomas completos esporádicamente), una alternativa mucho más económica es la externalización del servicio de computación mediante la utilización de servicios privados de computación deslocalizada (*cloud computing*).

Estos servicios constituyen una alternativa muy atractiva, ya que la facturación es por tiempo de computación utilizado, y no es necesario adquirir y mantener grandes infraestructuras. Los programas de análisis pueden ser instalados por bioinformáticos de la propia institución o pueden utilizarse paquetes de análisis de proveedores de servicios a terceros, que analicen los datos en la nube y generen los informes. Los mecanismos de acceso controlado de los profesionales médicos a esta información serán distintos de si se establece un sistema local adaptado a las necesidades del centro, pero es posible establecer diferentes usuarios con distintos privilegios de acceso para la consulta o análisis de este tipo de datos.

Los costes económicos del uso de tiempo de computación, análisis de datos y almacenamiento, y su comparación con una instalación local, deberán ser evaluados con objeto de tomar la decisión más adecuada para cada centro. Sin embargo, los términos de servicio de dichos proveedores, y el hecho de que la mayor parte de servidores estén localizados fuera del territorio nacional, no siempre se adecuan al tipo de protección que la legislación requiere para los datos genómicos, y lo que es más importante, la mayoría de los profesionales que tienen que decidir sobre las características técnicas de un servicio de este tipo no disponen del conocimiento ni asesoramiento legal necesario para decidir si alguna de las cláusulas de las condiciones de

servicio no se ajustan a la legislación, ni cómo subsanar dichas deficiencias. Otro aspecto que debe contemplarse en el caso de servidores comerciales es que las condiciones de servicio pueden modificarse unilateralmente. En caso de que dichas condiciones ya no se ajusten a la legislación nacional, o ya no sean beneficiosas para el centro hospitalario, la institución debería buscar alternativas para el almacenamiento de los datos genómicos asistenciales archivados en dicho proveedor de servicios (v. pág 59).

En cualquier caso, dependiendo de las necesidades que se generen dentro del Sistema Nacional de Salud, no sería descartable el desarrollo de infraestructuras de *cloud computing* públicos, similares a las redes de biobancos, que permitan dar soporte computacional asegurando la confidencialidad de los datos que se manejan.

Archivo y niveles de acceso a la información en investigación

Como se ha mencionado anteriormente, las necesidades de procesamiento y almacenamiento en investigación pueden ser muy diferentes a las exigidas para uso asistencial. En cualquier caso, los modelos organizativos son similares en ambos casos.

A diferencia de los datos de tipo asistencial, en los que siempre hay una conexión directa entre los datos genómicos y el sujeto fuente, en cualquier proyecto de investigación que implique la utilización de datos genómicos se realiza una codificación de los pacientes, normalmente previa a la toma de la muestra biológica de la que se va a obtener la secuencia. De esta manera, cada fichero genómico va asociado a un código que solo el médico del hospital participante puede saber a qué historial clínico corresponde, y no los investigadores. En el caso de los estudios familiares, los investigadores tienen más información porque es imprescindible conocer las relaciones de parentesco entre los individuos estudiados, aunque esto no implica que accedan a información sobre la identidad de los sujetos o sobre sus historias clínicas puesto que gestionan los datos también a través de códigos.

En cuanto a los modelos organizativos, son muy similares a los requeridos a nivel asistencial, con la diferencia de que la dimensión de las infraestructuras vendrá determinada por el tipo de proyecto y el número de grupos de investigación implicados en la misma. Los principales modelos organizativos son:

- **Servidores locales.** El uso de servidores locales suele ser la opción más extendida, debido a que la novedad de las técnicas de NGS hace que el número de grupos de investigación con acceso a datos de genomas completos o exomas sea todavía limitado, y en la mayor parte de los casos, los grupos suelen contar con capacidad de análisis al disponer de infraestructura para computación intensiva. Los servidores locales se pueden a su vez dividir en dos grupos:
 - **Servidores propios.** Son equipos cuyo uso está dedicado solo al análisis de este tipo de datos genómicos y por lo general se encuentran dentro del mismo laboratorio o institución a la que pertenece el grupo de investigación. El acceso a los mismos está restringido a los miembros del equipo investigador implicados en el análisis, y el acceso a los datos se suele llevar a cabo mediante conexiones encriptadas y protegidas mediante contraseñas seguras o claves de encriptación (tipo RSA).
 - **Servidores compartidos.** En la mayoría de los casos esta es la opción más utilizada, ya que desde el punto de vista institucional (universidad/centro de investigación) permite una

optimización de los recursos computacionales. Así, en muchos casos es frecuente que el grupo de investigación no cuente con un equipo informático dedicado a esta tarea, sino que la institución o el departamento tenga una infraestructura común de computación. En este caso, los miembros del equipo investigador tienen cuentas protegidas por contraseña, y la información genómica se almacena en el equipo, pero se necesitan privilegios especiales para acceder a esta información, privilegios que tienen los miembros del equipo y no el resto de usuarios de dicha infraestructura. Un caso especial de este tipo de modelo sería la utilización de infraestructuras de computación, nacionales, autonómicas o locales. A diferencia de los anteriores, los equipos no pertenecen a la institución en la que se realiza la investigación, sino a una red o consorcio público que proporciona acceso a recursos de computación para investigación. El acceso a dichos recursos suele estar regulado mediante convocatorias públicas y el establecimiento de convenios entre la institución investigadora y la que proporciona la infraestructura de computación, y los datos solo son accesibles a los miembros del equipo investigador.

- **Proveedores de servicios y servidores locales.** Al igual que en el caso de uso asistencial, el reciente desarrollo de las técnicas de secuenciación masiva hace que el número de profesionales con conocimientos suficientes para el análisis de este tipo de datos sea todavía limitado. No obstante, el espectacular abaratamiento de los costes de investigación ha permitido que muchos grupos de investigación biomédica puedan beneficiarse del empleo de estas nuevas herramientas. El hecho de que la mayoría de estos grupos no cuenten con personal cualificado para el análisis de este tipo de datos puede solventarse mediante la contratación de proveedores de servicios especializados en este tipo de análisis. Estos proveedores de servicio analizan las secuencias crudas para generar informes que luego se emplearán en la investigación. Una vez realizado el análisis, las secuencias son almacenadas por el grupo de investigación, bien en servidores locales, bien en simples discos duros si no se piensa reanalizar los datos.

De manera similar al uso asistencial, los términos de servicio deben regular los criterios de confidencialidad de los datos, así como el tiempo que los datos permanecen en los servidores del proveedor.

- **Servidores deslocalizados.** Por último, los sistemas deslocalizados (o *cloud computing*) constituyen una alternativa que permite acceder a poder de procesamiento y/o almacenamiento, sin necesidad de disponer de infraestructura computacional, o como recurso en periodos en que la demanda computacional del grupo o instituto es superior a la capacidad de este. Las ventajas de dichos servicios se han mencionado con anterioridad, y son fundamentalmente la facturación por tiempo de procesamiento utilizado o por cantidad de información almacenada y tiempo, no como en las infraestructuras tradicionales. Pero las condiciones de los términos de servicio, y el cumplimiento con la legislación vigente al depositar datos genómicos en servidores localizados en terceros países, o deslocalizados, constituyen riesgos que la mayor parte del personal investigador no tiene formación para valorar, necesitándose el apoyo de expertos en derecho para evaluar dichos términos de servicio.

Por otra parte, en el caso de investigación básica, este tipo de sistemas también tiene la ventaja de facilitar colaboraciones en las que se comparte gran volumen de datos. Un ejemplo son los proyectos PanCancer de secuenciación de genomas del cáncer (ICGC y TCGA), en los que se están analizando decenas de miles de genomas tumorales, equivalentes a más de 2 petabytes de datos. La transferencia y almacenamiento de este volumen de datos constituiría una tarea casi imposible para la mayoría de institutos de investigación, así

como un derroche de recursos el mantener copias locales de este volumen de datos. La utilización de servicios de *cloud computing* permite que todos los investigadores interesados en el análisis de estos datos puedan tener acceso a ellos en un mismo servidor, y que sean facturados dependiendo de los recursos computacionales que utilicen (tiempo de procesador, memoria, almacenamiento,...). No obstante, la solución adoptada por el TCGA estadounidense para el acceso a los datos genómicos ha sido la creación del Cancer Genomics Hub (CGHub), administrado por un instituto público como es la Universidad de California en Santa Cruz (UCSC). Por lo tanto, a la hora de utilizar servicios de *cloud computing* para el depósito y manejo de datos genómicos para investigación, deberá considerarse si se trata de un servicio público, creado para facilitar el análisis de este tipo de datos de acuerdo a la legislación correspondiente, o por el contrario, si se trata de servicios comerciales será necesario examinar los términos de servicio para evitar posibles conflictos con la legislación aplicable y garantizar la confidencialidad de los datos.

Papel potencial de los biobancos en la centralización, archivo y custodia de los datos

Los biobancos son considerados como las herramientas para conseguir que la investigación traslacional sea una realidad, puesto que son puentes entre la investigación básica y la clínica. Estos servicios tienen como objeto recabar, procesar y conservar muestras biológicas de diversa naturaleza en unas determinadas condiciones de calidad (biológica y de información asociada), y actuar como gestores en la cesión y envío de estas, para su uso posterior en investigación, y sus datos asociados, siempre bajo requisitos éticos y legales. Por tanto, cabe esperar que las actuales plataformas de secuenciación que se están erigiendo en centros hospitalarios e institutos de investigación puedan trabajar en estrecha colaboración con los que son los proveedores de muestras que utilizarán en sus análisis, los biobancos. A su vez, los biobancos precisarán de los datos obtenidos de estos análisis para “enriquecer” y dar valor añadido a las muestras conservadas. Independientemente del tipo de archivo, podría ser conveniente centralizar los datos obtenidos en NGS en servicios o unidades regulados convenientemente y trabajando con estándares éticos. Tal vez las unidades más adecuadas para ello sean los biobancos, puesto que todas sus actividades se realizan bajo unas determinadas condiciones de calidad, tanto de las muestras como de los datos asociados. En este sentido, los biobancos tendrían que establecer un plan de actuación que pudiese contemplar el ampliar el valor añadido de sus colecciones con los datos genómicos de las muestras relacionados.

Hoy por hoy, esta no es la realidad de los biobancos españoles: en general, no se almacenan datos genómicos junto con las muestras humanas registradas. Debe tenerse en cuenta que este tipo de ampliación del alcance de los biobancos llevaría asociado un reto de gestión²⁶, de coste²⁷ e incluso ético²⁸ como ha sido puesto de manifiesto por la comunidad científica internacional en los últimos años.

RESUMEN DEL CAPÍTULO

Los instrumentos de secuenciación generan miles o miles de millones de secuencias o lecturas. Durante el proceso de secuenciación, el secuenciador suele generar unos datos crudos que en algunas de las tecnologías más empleadas se corresponden con simples imágenes de microscopía. Estos datos crudos tienen que ser procesados, generalmente por el mismo equipo de secuenciación, y al final se generan los datos de secuencias crudas y sin procesar. Existen los siguientes formatos de archivos que son de uso habitual en NGS:

- a) **formato FASTQ:** es el formato más estándar y más empleado. En este formato cada lectura está representada por cuatro líneas;
- b) **formato BAM:** es el procedimiento para alinear frente al genoma de referencia, cada una de las lecturas, cuyo resultado se almacena en un formato específico que es un fichero SAM/BAM. Se almacenan todas las lecturas que estaban en el formato FASTQ, pero para cada lectura se anota en qué cromosoma y posición alinea mejor, con un parámetro de confianza en dicho alineamiento. El formato BAM almacena todas las lecturas alineadas contra el genoma de referencia, por lo que contiene la información necesaria para llevar a cabo los distintos tipos de análisis que se pueden realizar. El tipo de análisis más frecuente, encaminado a la identificación de variantes génicas causantes de una determinada patología, suele requerir al menos tres niveles de análisis:
 - 1) extracción de variantes genómicas,
 - 2) filtrado de variantes y
 - 3) identificación de variantes funcionales.
- c) **formato VCF:** es el formato más empleado para almacenar la información de variantes genéticas (VCF: *Variant Calling Format*). Este formato permite describir de manera sencilla la mayor parte de las variantes génicas, proporcionando información acerca del locus genómico en el que se localiza la variante (cromosoma y posición), qué secuencia es la presente en el genoma de referencia y cuál es la que se ha observado en el genoma analizado, y luego otros parámetros como el genotipo, y una puntuación que refleja la confianza que se tiene en la variante.

El parámetro más crítico para poder identificar cualquier variante génica es la cobertura. Este parámetro refleja cuántas lecturas se han obtenido correspondientes a la región de interés. Para cada variante identificada es necesario establecer algún parámetro probabilístico que refleje la confianza estadística que se tiene en dicha afirmación. El formato VCF permite describir una variante incluyendo toda esta información adicional, así como el tipo de filtros que se ha utilizado posteriormente, por lo que es el formato en el que se suelen representar todos los datos de variantes procesadas. Sin embargo, el número de variantes que pueden identificarse al secuenciar un exoma o un genoma es muy elevado, y la mayoría de estas variantes no tienen un efecto causal en la patogénesis de la enfermedad estudiada. Por esta razón, una vez identificadas todas las variantes, es necesario llevar a cabo un filtrado de los datos para reducir el número total de variantes, eliminando las no informativas, y almacenando las informativas. En cualquier caso, los ficheros generados, tanto con los datos crudos como con los filtrados, no contienen información clínica acerca del paciente secuenciado. No obstante, los datos crudos y las variantes sin filtrar contienen toda la información genómica de una persona, lo que podría permitir su identificación si los datos genéticos de dicha persona, o de algún familiar cercano, están depositados en otra base de datos.

A pesar de los progresos realizados en biología molecular y genética/genómica, el número de variantes de utilidad clínica o accionables, entendida la utilidad clínica como que el conocimiento de poseer una mutación concreta puede conducir a pautas de tratamiento específicas o al empleo de fármacos selectivos, es todavía muy pequeño. En el caso de cáncer puede decirse que en el mejor de los casos se han caracterizado unas decenas de mutaciones de las que se pueden deducir posibles pautas clínicas. Así mismo disponemos solo de unas decenas de fármacos aprobados para su utilización y sujetos a una regulación sobre casos de uso muy estricta.

No todas las variantes potencialmente identificadas tienen la misma implicación y relevancia para la salud de un individuo. Así, en un estudio genómico podemos encontrar:

- a) mutaciones que confirman la enfermedad que provocó el análisis del paciente;
- b) mutaciones que son clínicamente relevantes y cuyo conocimiento supone un beneficio directo para el paciente o su descendencia (ya sea porque se puede intervenir sobre ellas o bien porque, aunque no tienen tratamiento, su conocimiento puede ayudar al paciente a tomar decisiones reproductivas);
- c) mutaciones de alta penetrancia relacionadas con enfermedades de aparición tardía (con y sin tratamiento en la actualidad);
- d) mutaciones que no causan enfermedad en el paciente pero que pueden causarla a su descendencia (portadores de enfermedades autosómicas recesivas o portadoras de enfermedades ligadas al cromosoma X), y
- e) mutaciones que confieren una cierta susceptibilidad a padecer una determinada enfermedad compleja, o variantes y cambios relacionados con caracteres de base genética de dudosa relación con la salud humana.

No existe una pauta unificada acerca de la información que debe incluirse en los informes que resultan de un análisis genómico, ni tampoco en cuanto a los hallazgos inesperados que pueden llegar a producirse con este tipo de tecnologías. Los informes de un laboratorio clínico deben seguir unas normas para garantizar la calidad del resultado, para que el profesional solicitante y el paciente puedan entender el resultado y su fiabilidad, y para permitir la transmisión de información a través del consejo genético y dirigir la atención médica especializada.

Deben constar en los informes, y hacerlo de forma comprensible para el receptor, principalmente el profesional solicitante y también el paciente, los siguientes datos:

- a) el motivo clínico de la solicitud que permita dar respuesta al solicitante a una pregunta concreta. ¿Tiene el paciente la mutación A que confirmaría su enfermedad y permitiría utilizarla para el consejo genético?;
- b) el tipo de estudio: diagnóstico, predictivo, de portador, prenatal, de informatividad, farmacogenético, etc.;
- c) la metodología empleada, su sensibilidad y especificidad, la cobertura, y resto de datos que permitan conocer la calidad (en términos de validez analítica, validez clínica) y la fiabilidad de la información. En los resultados debe constar:
 - todos los datos relevantes del análisis de la secuencia estratificando su relevancia clínica: hallazgo de mutación para la enfermedad sospechada clínicamente, hallazgo de mutaciones patogénicas causantes de otras enfermedades con carácter de presentación

inmediata, predictivo y/o presintomático, hallazgo de variantes inciertas de las que no se conoce su carácter de variación normal y que por su posición genómica y la predicción de la función del gen concreto o por niveles de evidencia bajos (estudios animales, celulares) que permiten sospechar patogenicidad, y que deban ser reportadas;

- la interpretación de estos resultados en términos de relevancia clínica: la mutación identificada que confirma la enfermedad, otras mutaciones que pueden causar una enfermedad con mayor o menor probabilidad, abordando los conceptos de penetrancia, susceptibilidad, variabilidad en la expresión e incertidumbre, un cambio incierto puede hacer sospechar un riesgo de salud potencial y es tributario de seguimiento clínico;
- la implicación de los resultados en términos de manejo médico y de consejo genético. Existen dos modelos posibles: la referencia a la necesidad de seguimiento médico y consejo genético sin más explicaciones o la explicación que incluya elementos detallados de las implicaciones del resultado, con referencia a las fuentes que han generado la evidencia, tanto sea para el pronóstico y tratamiento del paciente, la forma de herencia y los riesgos asociados a la mutación como para un posible diagnóstico prenatal o preimplantacional en términos de opciones reproductivas disponibles, o para otros miembros de la familia con riesgo de tener la misma variante y sus implicaciones clínicas.

Existen en la actualidad diversas situaciones en los biobancos en cuanto al tratamiento y almacenamiento de los datos asociados a las muestras; sin embargo, de forma generalizada, los biobancos españoles no almacenan los datos genómicos de las muestras que gestionan. Así, desde el punto de vista clínico, los análisis genómicos constituyen una prueba diagnóstica que ayuda en la toma de decisiones médicas. Al igual que la mayoría de pruebas realizadas, los datos de una prueba genética deben incorporarse a la historia clínica del paciente. No obstante, el volumen de datos generado por los estudios genómicos hace que el archivo y gestión de estos datos sea distinto al de otras pruebas diagnósticas, por lo que es necesario valorar diferentes modelos de gestión de estos.

Los informes clínicos deben detallar las implicaciones para el paciente concreto y su familia, de forma personalizada, y deben acompañarse de las sesiones de información necesarias para comunicar de forma eficiente las implicaciones de la información genética/genómica. Existen, sin embargo, características de la información genética que la hacen particularmente tributaria de confidencialidad y de respeto singular:

- a) los datos identifican estados de portador o enfermo “para toda la vida”, no problemas transitorios que serán simplemente un antecedente médico una vez superados;
- b) los datos identifican enfermos, pero también portadores sanos, estados de salud presintomáticos o predicciones de riesgos de enfermedad;
- c) los datos son generados para el paciente pero pueden tener implicaciones para el resto de familiares;
- d) la complejidad de la información a transmitir con frecuencia no se corresponde con la formación en genética del profesional solicitante de la prueba. Por tanto, la gestión de los sistemas de información de los servicios de salud deberá adaptarse para garantizar el cumplimiento de la legislación vigente y la de futuro desarrollo, y deberá abordar una serie de aspectos importantes que actualmente no están resueltos, y hacerlo de forma común para los distintos servicios de salud de un país y probablemente de la Unión Europea.

La Ley 41/2002, de Autonomía del Paciente, representa un marco general que puede dar respuesta a estos puntos, puesto que prevé la obligación general de que toda la documentación clínica se archive en la historia y de que los profesionales accedan a los datos cuando sea pertinente por razones asistenciales. El archivo de datos genómicos procedentes de pruebas de diagnóstico asistencial requiere una planificación rigurosa debido al volumen de datos que se generan (hasta 100 gigabytes por genoma completo), a la demanda prevista para dichos análisis a medio y largo plazo, y a la necesidad de garantizar la privacidad de los datos. Por esta razón, y a pesar de que los análisis genómicos pueden ser solicitados por profesionales pertenecientes a distintos servicios, es aconsejable que el archivo de los datos se lleve a cabo de manera centralizada. Las necesidades de procesamiento y almacenamiento de esta información dependerán de:

- a) tipo de análisis utilizado;
- b) número de pacientes esperado, y
- c) tiempo de diagnóstico.

Teniendo en cuenta estos tres parámetros, y asumiendo la utilización de un mismo análisis bioinformático, será posible determinar los distintos escenarios y los requerimientos computacionales necesarios para conseguir los objetivos establecidos. En cualquier caso, y debido al previsible aumento en la demanda de este tipo de análisis, será necesaria la utilización de sistemas escalables que permitan adaptarse progresivamente a la demanda. El archivo de datos podrá ser realizado en:

- a) servidores locales (análisis y almacenamiento en servidores propios o análisis en servidores ajenos y almacenamiento en servidores propios o
- b) archivo en la nube y modelos de almacenamiento remoto.



MARCO JURÍDICO

LA IDENTIFICABILIDAD DEL SUJETO COMO CRITERIO PARA LA CALIFICACIÓN JURÍDICA DE LOS DATOS

El proceso de tratamiento de datos genéticos en el ámbito biomédico, que abarca todas las operaciones desde su obtención hasta su cancelación, ha sido objeto de una regulación específica en la Ley 14/2007, de 3 de julio, de Investigación Biomédica (LIB) y en el Real Decreto 1716/2011 de 18 de noviembre, por el que se establecen los requisitos básicos de autorización y funcionamiento de los biobancos con fines de investigación biomédica y del tratamiento de las muestras biológicas de origen humano, y se regula el funcionamiento y organización del Registro Nacional de Biobancos para investigación biomédica. Con carácter supletorio, en lo no previsto en la LIB se aplicará la legislación sobre protección de datos (Ley Orgánica 15/1999, de 13 de diciembre, de Protección de Datos de Carácter Personal [LOPD] y Real Decreto 1720/2007, de 21 de diciembre, por el que se aprueba el Reglamento de desarrollo de la Ley Orgánica 15/1999, de 13 de diciembre, de protección de datos de carácter personal [RLOPD]) y la legislación sobre derechos de los pacientes y documentación clínica (Ley 41/2002, de 14 de noviembre, básica reguladora de la autonomía del paciente y de derechos y obligaciones en materia de información y documentación clínica y legislación de desarrollo en las comunidades autónomas).

En concreto, el capítulo II del Título V de la LIB (“Análisis genéticos y tratamiento de datos genéticos de carácter personal”) es directamente aplicable al tratamiento de los datos genéticos, mientras que los capítulos III (“Utilización de muestras biológicas humanas con fines de investigación biomédica”) y IV (“Biobancos”) se refieren a la gestión de las muestras, aunque no explícitamente a la utilización posterior de los datos que se obtengan de los análisis, ni tampoco a los datos obtenidos a través de análisis de muestras “clínicas” (obtenidas y almacenadas con fines diagnósticos o terapéuticos, pero no de investigación). Sin embargo, en el desarrollo normativo de estos capítulos se ha definido el tratamiento de muestras de origen humano como “el conjunto de operaciones y procedimientos que permiten la obtención, conservación, almacenamiento, utilización y cesión de muestras biológicas de origen humano y, en su caso, de los datos asociados a las mismas” (artículo 2 n del RD 1716/2011). Por consiguiente, para la obtención, conservación, almacenamiento, utilización y cesión de los datos se habrá de observar lo dispuesto en la normativa que regula estas operaciones cuando tienen por objeto muestras biológicas.

Este panorama se completa con textos internacionales de distinta naturaleza y fuerza vinculante que han inspirado la regulación y pueden ser útiles cuando surjan dudas sobre su interpretación. Son destacables: la Declaración Universal sobre los Datos Genéticos Humanos de la UNESCO, de 16 de octubre de 2003; la Recomendación 4 del Consejo de Europa, sobre investigación con material biológico de origen humano, de 15 de marzo de 2006, y el Protocolo Adicional al Convenio de Biomedicina del Consejo de Europa sobre análisis genéticos clínicos, de 2008 (todavía no firmado por España).

Independientemente del contexto en el que se obtengan, utilicen, almacenen o cedan, la identificabilidad del sujeto de los datos es una condición determinante para establecer cuál es el régimen jurídico aplicable. En última instancia, lo que se pretende es que el sujeto mantenga el control de la información que puede ser relacionada con su identidad (cuándo se puede obtener, para qué y por quién). Como se verá, cuando esta posibilidad exista directa o indirectamente, los datos se consideran de carácter personal porque su tratamiento afecta, de hecho o potencialmente, a los intereses del sujeto. La clave, por tanto, es la posibilidad de “trazabilidad” desde la información hasta un individuo concreto.

El tratamiento de los datos genéticos se realiza en ocasiones con una asociación directa a identificadores (p. ej., cuando se reflejan en un informe clínico en el que figura el nombre y apellidos del paciente); en otros casos se manejan después de haber eliminado los identificadores (p. ej., en investigaciones sin beneficio directo para los sujetos y en las que no interesa realizar un seguimiento clínico); por último, con frecuencia, se archivan conservando un código que no es accesible para un investigador o para el laboratorio que realiza el análisis diagnóstico pero sí para el centro o profesional que los transfirió o que envió la muestra (se acude a esta última opción cuando interesa mantener una trazabilidad entre los datos y los pacientes, aunque el usuario no necesita identificarlos directamente). Los datos entrarán en una u otra categoría jurídica según las siguientes definiciones:

- **Datos de carácter personal:** “Cualquier información numérica, alfabética, gráfica, fotográfica, acústica o de cualquier otro tipo concerniente a personas físicas identificadas o identificables” (artículo 5.1.f del RLOPD). Por consiguiente son datos de carácter personal los resultados de cualquier análisis si el sujeto es identificable, independientemente de su nivel de interpretación (datos crudos de secuencias o identificación de variantes patológicas).
- **Datos de carácter personal relacionados con la salud:** “las informaciones concernientes a la salud pasada, presente y futura, física o mental, de un individuo. En particular, se consideran datos relacionados con la salud de las personas los referidos a su porcentaje de discapacidad y a su información genética” (art. 5.1.g del RLOPD).
- **Dato genético de carácter personal:** “información sobre las características hereditarias de una persona, identificada o identificable, obtenida por análisis de ácidos nucleicos u otros análisis científicos” (art. 3.j de la LIB).
- **Dato anónimo:** “dato registrado sin un nexo con una persona identificada o identificable” (art. 3.h de la LIB).
- **Dato anonimizado o irreversiblemente disociado:** “dato que no puede asociarse a una persona identificada o identificable por haberse destruido el nexo con toda información que identifique al sujeto, o porque dicha asociación exige un esfuerzo no razonable, entendiéndose por tal el empleo de una cantidad de tiempo, gastos y trabajo desproporcionados” (art. 3.i LIB).

- **Dato codificado o reversiblemente disociado:** “dato no asociado a una persona identificada o identificable por haberse sustituido o desligado la información que identifica a esa persona utilizando un código que permita la operación inversa” (art. 3.k de la LIB).

Como se ve, la razonabilidad en el esfuerzo para la identificación y la posibilidad de trazabilidad son criterios determinantes a la hora de distinguir estas categorías. A continuación se analiza el significado de cada uno de ellos:

- a) En cuanto a la razonabilidad en el esfuerzo que pueda dar lugar a la identificación del sujeto de los datos, el Dictamen 4/2007 sobre el concepto de datos personales, del Grupo de trabajo del artículo 29 de la Directiva 95/46/CE, señala que se deben tener en cuenta “la finalidad del tratamiento, la manera en que el tratamiento está estructurado, el rédito que espera obtener el responsable del tratamiento, los intereses individuales en juego, así como el riesgo de que se produzcan disfunciones organizativas (p. ej., un quebrantamiento del deber de confidencialidad) y los fracasos técnicos”; en cualquier caso, “el test de la razonabilidad de la identificabilidad” debe ser dinámico y tener en cuenta el contexto general y el conjunto de los datos disponibles por la comunidad científica.

Según aquel dictamen: “En los casos en que la identificación del interesado no entre en la finalidad del tratamiento, las medidas técnicas para evitar la identificación tienen un papel muy importante. Adoptar las medidas técnicas y organizativas adecuadas de acuerdo con los conocimientos técnicos existentes para proteger los datos contra la identificación puede constituir la clave para considerar que las personas no son identificables, teniendo en cuenta el conjunto de medios que puedan ser razonablemente utilizados por el responsable del tratamiento o por cualquier otra persona, para identificarlas. En este caso, la aplicación de esas medidas no es consecuencia de una obligación jurídica impuesta por el artículo 17 de la Directiva (que únicamente se aplica si la información se considera como datos personales a primera vista), sino una condición para que precisamente la información no se considere como datos personales y su tratamiento no esté sujeto a la Directiva”.

Antes ya se explicó que los ficheros generados de los análisis genéticos, tanto con los datos crudos como con los filtrados, contienen toda la información genómica de una persona, lo que podría permitir su identificabilidad si los datos genéticos de dicha persona, o de algún familiar cercano, están depositados en otra base de datos “indubitada” (relacionada con el sujeto identificado). Es decir, se plantea que esas secuencias podrían ser consideradas como identificadores.

Pues bien, teniendo en cuenta el criterio de razonabilidad mencionado anteriormente como parámetro para determinar si estos datos son identificadores y, por consiguiente, deben ser tratados como datos de carácter personal, habrá que valorar las posibilidades que existen en relación con la identificabilidad y las medidas organizativas que se hayan tomado para evitar dicha identificación. Esto significa que el contexto en el que se traten los datos y las condiciones que revista este tratamiento son circunstancias que pueden determinar si los datos deben ser considerados como de carácter personal o no.

Para evitar la identificación de los sujetos cuyos datos genéticos están almacenados en estas bases de datos, se podrían adoptar las siguientes medidas:

- Precaución en relación con la finalidad, que se puede concretar en el control y autorización para cada acceso a las bases de datos genéticos potencialmente identificadores, de manera que solo se permita para investigaciones aprobadas, con interés científico justificado, y en la aplicación de medidas de seguridad que permitan el control auditable de dichos accesos.
- Precaución en relación con la futura utilización por parte del destinatario, que se puede concretar en la firma de un compromiso legalmente vinculante entre las partes en el que figuren las condiciones del tratamiento: compromiso de confidencialidad y no identificación de los sujetos, y utilización para el fin concreto autorizado. En el compromiso se deberían incluir las consecuencias de su incumplimiento, que se podrían referir a: suspensión del acceso y veto para otros posteriores; comunicación de la infracción a quien corresponda si se pretende la publicación de resultados de la investigación llevada a cabo en términos ilícitos, y/o sanciones económicas.
- Establecer un sistema de control externo al investigador que podría corresponder al CEI o a la dirección del centro, que deben estar involucrados en el acuerdo mencionado.
- Precaución en relación con la transferencia y el almacenamiento de los datos por parte del destinatario, que se podría concretar en la exigencia de medidas de seguridad proporcionales.
- Precaución en relación con el respeto a los derechos del sujeto fuente, que se podrían concretar en la necesidad de información y consentimiento en relación con la anonimización y el futuro uso de los datos.
- Precaución en relación con el mantenimiento de las condiciones en el futuro, que se podrían concretar con medidas de transparencia (publicación y accesibilidad a las políticas de gestión de los datos) y previsiones a largo plazo (identificación de las obligaciones del responsable de los datos). En relación con este último punto se debe llamar la atención sobre la necesidad de que se adquieran compromisos de seguimiento de la gestión de los datos por parte de la institución del investigador que, más en concreto, podrían recaer sobre el CEI.

Es importante señalar que estas medidas deben ser proporcionales a las posibilidades de identificación según el contenido de las bases de datos (variantes génicas, mutaciones patológicas, mutaciones somáticas) y deben tener carácter dinámico, es decir, adaptables a la evolución de la técnica y al contexto en que se manejen.

- b) Por lo que se refiere a la trazabilidad, se debe señalar que puede ocurrir que los datos en sí mismos no sean directa o indirectamente identificativos por parte de quien los maneja pero que, de hecho, se mantenga la trazabilidad con el donante o paciente. En efecto, como se dijo en la página 46 y siguientes, lo más frecuente es que al inicio de cualquier proyecto de investigación que implique la utilización de datos genómicos, o cuando se trasladan datos clínicos a registros no clínicos (bases de datos de investigación, registros de enfermedades, registros epidemiológicos, etc.) se proceda a la “codificación de los pacientes”, de tal forma que cada fichero genómico vaya asociado a un código que se conserva en poder del centro donde se trata al paciente o donde se donó la muestra. Solo la persona autorizada en ese centro puede saber a qué sujeto corresponde la infor-

mación. Pues bien, la cesión de estos datos codificados, aun cuando quien los maneja no tenga acceso a la identidad de los pacientes, requiere también el consentimiento del paciente, puesto que del mantenimiento de la trazabilidad derivan posibles incidencias en su derecho al control de sus datos (piénsese, por ejemplo, en la posibilidad de hallazgos inesperados).

No obstante, pueden surgir dudas en relación con la aplicación a estas bases de datos codificados de todo el régimen del tratamiento de datos de carácter personal: la lógica de la aplicación del régimen del tratamiento de datos personales de la LOPD a una base de datos codificados debe justificarse con criterios de razonabilidad. ¿Con qué finalidad se da de alta esta base de datos codificados en la Agencia de Protección de Datos? ¿Podrá el responsable del tratamiento –el investigador o su institución– cumplir con sus obligaciones como tal y garantizar los derechos de acceso y cancelación del sujeto de los datos? En definitiva, más allá de lo previsto para estos datos en la LIB, ¿se debe aplicar el régimen de la LOPD para datos disociados?

En el Código Tipo de Farmaindustria se ha optado por responder negativamente a estas preguntas en el caso de los datos codificados resultantes de los ensayos clínicos que se almacenan en las bases de datos gestionadas por el promotor si se adoptan determinadas garantías jurídicas vinculantes (lo que podría equivaler a la firma del compromiso antes mencionado) y organizativas (a las que también se ha aludido). En algunos informes de la Agencia Española de Protección de Datos (AEPD) también se sigue esta línea (p. ej., en el Informe 207/2008, relativo a la creación de una base de datos clínicos, patológicos, radiológicos, de seguimiento sin que en ella aparezca ninguna identificación personal del paciente y en la Opinión 06/2013 sobre reutilización de datos abiertos en el sector público). En relación con las bases de datos genéticos, esta cuestión no ha sido suficientemente debatida.

Como base para esta reflexión se propone tener en cuenta lo siguiente cuando exista trazabilidad pero la cesión se lleve a cabo de forma codificada:

- Se debe aplicar lo dispuesto en la LIB para los datos reversiblemente disociados y, por tanto, se ha de contar con el consentimiento del sujeto para la utilización de sus datos genéticos codificados con fines de investigación (arts. 50.2 y 3 de la LIB). Este consentimiento debe seguir las pautas previstas para la gestión de las muestras (ser específico para un proyecto, para una línea de investigación, o amplio para cualquier investigación biomédica en un biobanco. Aunque varios fines pueden presentarse simultáneamente en un mismo documento). A partir de aquí, se ha de tener en cuenta que el tratamiento de datos codificados puede afectar al derecho de control sobre la información. Las condiciones del tratamiento de los datos deben ser adecuadas para que se garantice este derecho de control y para gestionar los riesgos que en cada caso concreto se presenten. La posibilidad de hallazgos inesperados relevantes para la salud (v. pág. 78) a partir de datos reversiblemente disociados es un ejemplo de esta situación: el investigador que ha descubierto la mutación no puede identificar al sujeto, pero quizás sí podría contactar con el médico para que sea este quien le comunique dicho descubrimiento.
- Se deberían articular mecanismos para garantizar el ejercicio de los derechos de los sujetos, en relación con la base de datos codificados, a través de la institución que mantenga las claves para la identificación.

- Como se propuso para los datos genéticos considerados anonimizados, se deben seguir una serie de **precauciones** en la gestión de los datos codificados, que representen garantías jurídicas y organizativas para la protección del derecho fundamental a la protección de datos personales y para salvaguardar la seguridad de la información. Recordemos que se referían a: precaución en relación con la finalidad del tratamiento; precaución en relación con la futura utilización por parte del destinatario; establecimiento de un sistema de control externo al investigador; precaución en relación con la transferencia y el almacenamiento de los datos por parte del destinatario, que se podrían concretar en la exigencia de medidas de seguridad proporcionales; precaución en relación con el mantenimiento de las condiciones en el futuro.

Tanto en el caso de datos anonimizados (cuando, por ser determinados datos genéticos, potencialmente podrían considerarse identificables aunque no por medios razonables) como en el caso de datos codificados, se debería valorar la oportunidad de impulsar un sistema de acreditación y registro de centros para su gestión. Este sistema sería más interesante si tuviera carácter internacional. Esta acreditación debería representar ventajas en la evaluación ética y jurídica de los proyectos que dichos centros llevaran a cabo. Ya existen algunas iniciativas en el marco internacional en esta línea, como la constitución de la llamada *Global Alliance for genomics and health*. Este foro agrupa a centros de investigación de un buen número de países de varios continentes y su objetivo es acordar criterios de armonización para facilitar la colaboración científica internacional en el intercambio de datos. El *Regulatory and Ethics Working Group* ha publicado el *Framework for Responsible Sharing of Genomic and Health-Related Data*, al que están llamadas a adherirse las instituciones que manejen datos genéticos con fines de investigación. Es un documento con previsiones de carácter general que previsiblemente se desarrollará y concretará en el futuro y que puede tener un potencial interesante para evaluar la gestión de los centros. Los promotores, de hecho se basaron en el sistema de acreditación de los “puertos seguros”, propio de la legislación sobre protección de datos²⁹.

LOS PRINCIPIOS DE LA PROTECCIÓN DE LOS DATOS GENÉTICOS DE CARÁCTER PERSONAL

Señala el artículo 4 de la LOPD que los datos de carácter personal solo podrán ser objeto de tratamiento cuando sean “adecuados, pertinentes y no excesivos en relación con el ámbito y las finalidades determinadas, explícitas y legítimas para las que se hayan obtenido”. Por consiguiente, únicamente se debe proceder al tratamiento de datos personales si es necesario para un determinado fin legítimo. Los datos serán cancelados cuando hayan dejado de ser necesarios o pertinentes para la finalidad legítima para la cual hubieran sido recabados o registrados.

Como se ha descrito en apartados anteriores, la obtención y el almacenamiento de secuencias genéticas son necesarios para la investigación biomédica que permita avanzar hacia una medicina personalizada en beneficio de los ciudadanos. Puede afirmarse, por lo tanto, que la creación y el acceso a bases de datos genéticos son legítimos cuando persigan una finalidad

de investigación científicamente validada o de protección de la salud del paciente o de la salud pública. Por consiguiente:

- Las especiales características de las secuencias genéticas requieren el control de acceso para garantizar que se realiza con una finalidad legítima, dentro de las señaladas en la legislación.
- El periodo de conservación de los ficheros, como se especificará más adelante, puede ser muy dilatado, incluso indefinido.

La incorporación de los datos de un proyecto a bases de datos disponibles para la comunidad científica superará el periodo de su desarrollo: incluso agotada la muestra y con el proyecto finalizado, tendría sentido conservar los datos en bibliotecas de datos que sirvieran de consulta y comparación para estudios posteriores. Por otra parte, tal vez, sería arriesgado y poco responsable eliminar los datos obtenidos una vez finalizado el proyecto de investigación y lo más prudente sería conservarlos durante un tiempo, de manera similar a como se funciona con los archivos de historias clínicas, por si pudieran surgir nuevos planteamientos o estudios familiares. Ahora bien, el sujeto debe ser informado de las previsiones sobre la conservación de los datos y sus finalidades, y mantener la capacidad de decisión sobre estas, salvo que concurra una habilitación específica.

De hecho, con frecuencia los datos genéticos clínicos se guardan indefinidamente. Esta situación no queda descartada en el artículo 52 de la LIB, según el cual: "1. Los datos genéticos de carácter personal se conservarán durante un período mínimo de cinco años desde la fecha en que fueron obtenidos, transcurrido el cual el interesado podrá solicitar su cancelación. 2. Si no mediase solicitud del interesado, los datos se conservarán durante el plazo que sea necesario para preservar la salud de la persona de quien proceden o de terceros relacionados con ella"; pero probablemente habría que determinar cuántas generaciones (o en función de la extensión de la familia) debe estar un dato genómico "personalizado" o asociado a una familia. Se debería plantear cuándo el dato debe ser anonimizado de forma irreversible en el contexto clínico.

LOS SUJETOS IMPLICADOS. TITULAR, RESPONSABLE Y ENCARGADO DEL TRATAMIENTO

La identificación de las categorías jurídicas en las que encajan los sujetos e instituciones implicados en el tratamiento de los datos es fundamental para determinar cuáles son sus derechos y obligaciones y cómo deben formalizarse las relaciones entre ellos³⁰.

- a) El sujeto del que proviene la muestra que se ha analizado ("sujeto fuente", art. 3.v de la LIB) es el "afectado o interesado" (art. 5.1.a del RLOPD), esto es, el titular de los datos objeto de tratamiento. Exclusivamente a él corresponde el ejercicio de los derechos reconocidos en la LIB y en la LOPD en relación con los datos.

Sus familiares, en principio, no tienen la categoría de interesados o afectados, aunque pudieran compartir con el sujeto fuente la información genética (cosa que no se sabe con

certeza hasta que se les realicen a ellos los correspondientes análisis). No obstante, este hecho se tiene en cuenta también en la gestión de la información, como interés que pesa en la resolución de ciertos conflictos que pudieran surgir, pero sin que se justifique el reconocimiento de un correspondiente derecho subjetivo sobre la información. Por ejemplo, cuando el titular opta por no conocer los resultados de un análisis pero estos pueden ser relevantes para la salud de sus familiares, se articulan mecanismos para hacerles llegar a estos la información. Es el derecho a la protección de la salud y a la integridad física, y no el derecho de acceso a los propios datos, el título jurídico que justifica la comunicación de la información en estos casos, es decir, es el bien que se quiere proteger.

b) El responsable del fichero o del tratamiento es la “Persona física o jurídica, de naturaleza pública o privada, u órgano administrativo, que sólo o conjuntamente con otros decida sobre la finalidad, contenido y uso del tratamiento, aunque no lo realizase materialmente” (artículo 5.1 q del RLOPD). Cuando quien decida sobre estas cuestiones sea el profesional, será este quien asuma la función de responsable del tratamiento. Cuando el profesional desempeñe su actividad en el marco de una institución a la que correspondan aquellas decisiones, aunque sea de manera global, será dicha institución la que asuma la función de responsable del tratamiento. Los profesionales serán entonces usuarios autorizados con sistemas de autorización y autenticación personalizados. Los deberes como responsable corresponderán por consiguiente al titular del fichero del centro. El profesional trata los datos como usuario autorizado. Al responsable del fichero le corresponde:

- Notificar los ficheros ante el Registro General de Protección de Datos para que se proceda a su inscripción.
- Asegurarse de que los datos sean adecuados y veraces, obtenidos lícita y legítimamente y tratados de modo proporcional a la finalidad para la que fueron recabados.
- Garantizar el cumplimiento de los deberes de secreto y seguridad.
- Informar a los titulares de los datos personales en la recogida de estos.
- Obtener el consentimiento para el tratamiento de los datos personales.
- Facilitar y garantizar el ejercicio de los derechos de oposición al tratamiento, acceso, rectificación y cancelación.
- Asegurar que en sus relaciones con terceros que le presten servicios, que comporten el acceso a datos personales, se cumpla lo dispuesto en la LOPD.
- Cumplir, cuando proceda, con lo dispuesto en la legislación sectorial que le sea de aplicación.

c) El encargado del tratamiento. Es la persona “jurídica, pública o privada, u órgano administrativo que, solo o conjuntamente con otros, trate datos personales por cuenta del responsable del tratamiento o del responsable del fichero, como consecuencia de la existencia de una relación jurídica que le vincula con el mismo y delimita el ámbito de su actuación para la prestación de un servicio” (art. 5.1.i del RLOPD).

El acceso a los datos por parte de un encargado del tratamiento no se considera una cesión (art. 12.1 de la LOPD) siempre que la prestación de servicios esté regulada en un contrato escrito con un contenido mínimo que es el siguiente (art. 12.2 y 3 de la LOPD):

- Que el encargado del tratamiento únicamente trate los datos conforme a las instrucciones del responsable.

- Que no los utilice con un fin distinto del previsto en el contrato ni los comunique, ni siquiera para su conservación, a terceros.
- Que se estipulen las medidas de seguridad que debe implementar el prestador del servicio.
- Que, una vez cumplida la prestación del servicio, los datos se devuelvan al responsable o se destruyan.

Una posible modalidad de devolución de los datos es la cesión de estos a un nuevo prestador de servicios designado por el responsable y que haya suscrito con este un contrato con las garantías del artículo 12 de la LOPD.

En el caso de que para la prestación de servicios intervengan terceros subcontratistas deberán especificarse los servicios a subcontratar, comunicar al responsable la identidad de los subcontratistas y celebrar un contrato entre el encargado del tratamiento y los subcontratistas con las mismas garantías que se han descrito anteriormente.

Una de las principales modalidades de encargado del tratamiento es la prestación de servicios de *cloud computing*.

La contratación de servicios de *cloud computing* se ha convertido en una opción a veces imprescindible y, al menos, de gran utilidad para el tratamiento de datos en la investigación biomédica.

La prestación de dichos servicios ha supuesto un cambio de paradigma en la aplicación de los criterios tradicionales de la normativa de protección de datos personales que exige definir nuevos criterios para garantizar los derechos de los ciudadanos.

El punto de partida para la contratación de estos servicios se encuentra en la exigencia de una diligencia cualificada por parte de quienes la llevan a cabo para informarse sobre la naturaleza de estos y sobre las garantías exigibles para cumplir con la normativa antes citada.

El primer aspecto que hay que considerar es obtener información sobre la modalidad de prestación de estos servicios (nube pública, privada, híbrida, comunitaria,...), ya que cada una de ellas implica mayores o menores riesgos y, por tanto, mayores o menores garantías en su contratación, en particular, si implica subcontrataciones de servicios y transferencias internacionales de datos a países que carecen de un nivel adecuado de protección.

Debe tenerse en cuenta que la posición jurídica de los contratantes de estos servicios es la de responsable del fichero o tratamiento para quien los contrata, por lo que ha de ser diligente en la obtención de información sobre sus características y sobre las garantías que deben articularse. Y la de encargado del tratamiento por parte del prestador del servicio, por muy relevante que sea la entidad que los ofrece.

Esta posición jurídica implica que la ley aplicable al contrato será la del cliente que contrata los mencionados servicios, ley que no puede ser modificada por pactos entre las partes.

En la prestación de estos servicios es habitual que intervengan terceras empresas subcontratadas, en cuyo caso el cliente que los contrata debe obtener la información sobre esta posibilidad y obtener garantías sobre las empresas subcontratistas.

Además, las empresas que intervienen en la prestación de estos servicios (encargados del tratamiento o subcontratistas) pueden estar ubicadas en terceros países que no ofrezcan garantías adecuadas, por lo que es imprescindible obtener información sobre esta cuestión y sobre las que ofrecen para realizar transferencias internacionales de datos.

Los datos genéticos y las muestras biológicas son datos especialmente protegidos, por lo que las medidas de seguridad exigibles en la contratación de servicios de *cloud computing* son de nivel alto, lo que obliga a obtener información sobre ellas y sobre cómo se auditará su cumplimiento y sobre las incidencias de seguridad que puedan producirse.

Al tratarse de datos especialmente protegidos, deben exigirse garantías de **confidencialidad** por parte del prestador de estos servicios.

Dada la dependencia del cliente respecto de la entidad prestadora de servicios de *cloud computing* es imprescindible obtener información y garantías sobre cómo podrá recuperarse la información facilitada para la investigación biomédica (portabilidad).

También deben obtenerse garantías para impedir el acceso por terceros, incluidas autoridades públicas de terceros países, sobre los datos objeto de la investigación biomédica y sobre su conservación o destrucción en caso de que esta sea procedente.

Del mismo modo, si fuera necesario, deben preverse protocolos que permitan al cliente posibilitar el ejercicio de los derechos de acceso, rectificación, cancelación y oposición o sobre las consecuencias de la revocación del consentimiento por parte del sujeto fuente.

Para obtener información más detallada sobre las garantías que se han expuesto, puede acudir a la *Guía para clientes que contraten servicios de cloud computing*³¹.

Si los contratantes de estos servicios son administraciones públicas, deberán cumplir, también, las obligaciones del Esquema Nacional de Seguridad (Real Decreto 3/2010, de 8 de enero).

MEDIDAS DE SEGURIDAD DE LOS FICHEROS

Ni la LIB ni el RD 1716/2011, ni la Ley 41/2002 recogen previsiones concretas en relación con la seguridad de los ficheros. El régimen aplicable a esta cuestión es el previsto en la LOPD y en el RLOPD 1720/2007. Así, en el tratamiento de los datos genéticos de carácter personal (crudos, filtrados, informes clínicos), como datos de salud, se deben observar las medidas de nivel alto además de las de nivel básico y medio, recogidas en esta normativa (v. la *Guía de Seguridad de la Agencia Española de Protección de Datos*³²).

Las medidas de seguridad de nivel alto se refieren a los siguientes aspectos:

- Designación de un responsable de seguridad.
- Identificación y formación del personal.
- Registro de incidencias.
- Control de acceso.

- Identificación y autenticación.
- Gestión de soportes.
- Copias de respaldo.
- Criterios de archivo que permitan el ejercicio de los derechos ARCO.
- Auditoría externa.
- Transmisión de datos por redes de comunicaciones (encriptación).

El responsable del fichero tiene el deber de garantizar el cumplimiento de estas medidas, que deben constar en un documento interno de la organización, que debe mantenerse siempre actualizado. Este deber también corresponde al encargado del tratamiento.

Los apartados mínimos que debe incluir el documento de seguridad son los siguientes: ámbito de aplicación (especificación detallada de los recursos protegidos); medidas, normas, procedimientos, reglas y estándares de seguridad; funciones y obligaciones del personal; estructura y descripción de los ficheros y sistemas de información; procedimiento de notificación, gestión y respuesta ante incidencias; procedimiento de copias de respaldo y recuperación de datos, y medidas adoptadas en el transporte, destrucción y/o reutilización de soportes y documentos.

En particular, es preciso insistir en la necesidad de que se lleve a cabo y se analicen los resultados del control de accesos y de la auditoría de seguridad. Esta medida de seguridad es especialmente relevante porque permite conocer el adecuado cumplimiento de las medidas de seguridad y porque, en la experiencia de la AEPD, se ha constatado que es una medida con un nivel reducido de cumplimiento.

INTERVENCIÓN DE COMITÉS DE ÉTICA

Las implicaciones éticas y jurídicas planteadas por la NGS no son completamente nuevas en relación con las que ya se suscitaron ante la irrupción de la genética y la genómica. Sin embargo, las características específicas de la NGS, tales como la cantidad y la calidad de los datos y las oportunidades abiertas para futuros análisis y reanálisis de los datos, así como las nuevas posibilidades técnicas que se abren para su gestión, sugieren la especial relevancia de las políticas que se articulen para su tratamiento. Por ejemplo, como se ha dicho, un sistema de acuerdos legalmente vinculantes entre las partes implicadas, la involucración de los centros o la posibilidad de acreditaciones para agilizar las tramitaciones de accesos o cesiones, son herramientas sobre las que se debe profundizar. La revisión de su puesta en práctica para cada escenario concreto por parte de expertos completa este marco de actuación.

Por esta razón, la intervención de los comités de ética adquiere una gran relevancia en este contexto. A continuación se describe cuál es su papel y qué dudas se pueden suscitar en relación con sus funciones.

Comités de ética asistencial

Los comités de ética asistencial (CEA) son órganos multidisciplinares cuyo cometido esencial es asesorar para la toma de decisiones en situaciones que pueden conllevar conflictos de valores en aras a proteger a los pacientes y a los profesionales en contextos asistenciales. La interven-

ción de los comités en el contexto clínico no es obligatoria ni vinculante, pues sus recomendaciones pueden o no ser seguidas por el profesional responsable o el centro, a diferencia del contexto de investigación en que la intervención de los comités de ética de la investigación sí es vinculante, de modo que la apreciación de conflictos éticos irresolubles puede conllevar el veto a desarrollar un proyecto de investigación.

En el ámbito de los “datos genómicos”, el papel de los comités de ética será distinto de forma clara según hablemos de clínica o de investigación, y así viene establecido en la propia normativa aplicable. La LIB regula los comités de ética de la investigación, con carácter vinculante para la aprobación de proyectos o para el visto bueno a determinadas excepciones a la norma general (p. ej., cesión de muestras de los biobancos sin consentimiento informado), mientras que en el ámbito clínico serán los CEA quienes podrán tener un papel en ciertos conflictos relacionados con los datos genéticos.

A pesar de que ni la LIB ni la LAP recogen mención alguna a los CEA, su tarea en clínica puede ser importante y relevante para esta casuística. El uso de la información genética en el contexto clínico puede generar conflictos de difícil gestión por parte del profesional y ser aconsejable recabar el consejo o recomendación del CEA.

Esta intervención puede darse en relación con la confidencialidad cuando se dé conflicto con datos de terceros que el sujeto afectado no quiera comunicar o no autorice la comunicación, previsión que ya recoge la propia LIB en caso de hallazgos inesperados cuando la persona ha manifestado previamente su voluntad de no ser informado. En este caso, el CEA podría valorar si prevalece el derecho a saber de terceros potenciales afectados frente al derecho a la confidencialidad del titular de esa información que ha renunciado a conocerla. También puede ser motivo de controversia en la que pueda intervenir el CEA la gestión de datos de menores cuando los padres insisten en diagnosticar y saber, o en la petición de datos de personas fallecidas (p. ej., acceso a la historia clínica), si el contenido de la información genética puede ser trascendente para familiares del fallecido. Otro escenario en el que el comité tendría un papel importante podría ser en casos de duda en el consejo genético, antes de trasladar esa información a los progenitores, a fin de ponderar la maleficencia/beneficencia de dicha información cuando hablamos de probabilidades o meros riesgos científicamente no demostrables.

Cada vez es más frecuente también que al profesional le soliciten una segunda opinión o interpretación respecto a informes o datos genéticos obtenidos de otras fuentes, incluso a partir de test o pruebas gestionadas por Internet y de escasa o nula fiabilidad, o bien la validación de informes realizados en otras instituciones. El profesional puede encontrarse en la duda de si debe o no acceder a esa valoración, puesto que según sea hacerlo o no hacerlo puede generar mayor daño al paciente –angustia, ansiedad–. Sin embargo, debe tenerse en cuenta que el acceso a la información genética forma parte ya de las prestaciones del Sistema Nacional de Salud, por lo que el ciudadano está en su derecho de recabar esa opinión.

Respecto a esta cuestión, en contextos como el de gestión de la información clínica que se lleva a cabo en Cataluña, donde se encuentra ya disponible y accesible al ciudadano su “carpetita personal de salud”, consultable telemáticamente, se debate la cuestión de si los datos genéticos deben también estar al alcance de su titular sin una adecuada interpretación. Esta es ya una cuestión que sobrepasa la posible consulta a un CEA y debe dirimirse a otro nivel

superior, como sería el comité de bioética de ámbito autonómico si lo hay, o en su defecto, poder hacer una consulta al Comité de Bioética de España. El conflicto ético-legal se daría entre el ejercicio de la autonomía y el derecho de acceso a sus datos, y el potencial daño de ese acceso directo sin la mediación-interpretación profesional adecuada. Por esta razón, puede ser razonable que cuando se solicite el acceso a datos genéticos, este derecho deba satisfacerse a través de vías específicas y más adecuadas que otras que funcionan para otros datos de la historia clínica (como, por ejemplo, que el servicio de atención al paciente entregue copia de un informe o análisis). La intervención de un profesional es recomendable en esta situación o, como mínimo, la advertencia al paciente de la trascendencia y significado de los datos que se le facilitan junto con la información sobre disponibilidad de consejo genético. El CEA puede cumplir una función de asesoramiento en este sentido.

En síntesis, como norma general y deseable, la gestión de la información genética debe dejarse en manos de los profesionales, formados y expertos, que puedan por sí mismos o dentro del equipo solucionar posibles conflictos con el paciente o sus familiares o terceros; pero en ocasiones el profesional puede tener dudas o no serle fácil encarar el problema en solitario, por lo que la intervención del CEA le puede suponer una ayuda importante. Aun así, en la práctica es poco habitual la intervención del CEA en el contexto clínico en relación con la gestión de datos genéticos y sus derivadas. Únicamente, y de modo excepcional, llegan a los comités consultas en este sentido por parte de los profesionales (p. ej., paciente donante de óvulos diagnosticada de Huntington, generando la dificultad de la trazabilidad de esos óvulos, o bien petición de padres de diagnóstico genético de sus hijos en relación con patología no tratable y de aparición tardía).

Si este tipo de consultas se extendiesen, una dificultad añadida sería la falta de formación de los miembros de los comités para analizar estas cuestiones. Sería pertinente que dicha información y conocimiento específico en genética se aportase “ad hoc” por parte de expertos y que el CEA, con dicho conocimiento, pudiese hacer la valoración ética del tema, emitiendo una recomendación para el profesional y/o el paciente, desde un análisis y visión más cercana al ciudadano que al experto.

Vistas las dificultades con las que hoy se trabaja en el campo genético, en cuanto a los controles y supervisión ética en el manejo de la información, y la celeridad con la que se desarrolla el conocimiento, sería deseable abordar soluciones a medio plazo, que podrían pasar por:

- Potenciar la función de los CEA, recogiendo en la normativa su papel en cuestiones éticas en el marco también de la gestión de los datos genéticos.
- Las instituciones y centros que trabajan con datos genéticos –en clínica y en investigación–, y que disponen de comités, deberían propiciar la relación entre ellos (CEI – CEI biobanco – CEA) a fin de actualizar periódicamente en qué se trabaja y cuáles son las cuestiones éticas implicadas.

Comités de ética en el contexto de la investigación

En España, aunque en 1978 se publicó un Real Decreto en el que se establecía la necesidad de que los ensayos clínicos con medicamentos fueran evaluados por el comité de ensayos clínicos del centro hospitalario, fue la Orden de 3 de agosto de 1982 del

Ministerio de Sanidad, el primer documento en el que se establecía la composición de estos comités, sus funciones, etc. Estas funciones se concretaron en la Ley 25/1990, del Medicamento, que estableció que los comités debían velar por el control ético, metodológico y jurídico de la investigación; y se desarrollaron en el Real Decreto 561/1993, que se refiere a los requisitos para la realización de ensayos clínicos con medicamentos. Más adelante, este fue sustituido por el Real Decreto 223/2004, de 6 de febrero, que traspuso la Directiva 2001/20/CE, de 4 de abril, que establece los criterios de evaluación ética de los ensayos clínicos y la composición y funcionamiento de los comités de ética de investigación clínica (CEIC).

En 2007, con la publicación de la LIB, se crean los CEI y se incluye en su ámbito de actuación la evaluación de toda investigación biomédica en seres humanos y la investigación con muestras biológicas incluida la obtención, donación y utilización de ovocitos, espermatozoides, preembriones, embriones, y fetos humanos o células, tejidos u órganos con fines de investigación, y no solo la referida a medicamentos y productos sanitarios.

De acuerdo con la LIB, los comités de ética de la investigación son órganos colegiados independientes y de composición multidisciplinar, cuya finalidad es la de velar por la protección de los derechos, seguridad y bienestar de los sujetos que participen en un proyecto de investigación biomédica y ofrecer garantía pública al respecto mediante un dictamen acerca de la documentación correspondiente del proyecto de investigación.

La LIB, en su Disposición transitoria tercera establece que “Los Comités Éticos de Investigación Clínica dejarán de existir a partir del momento en que se constituyan los Comités de Ética de la Investigación. Hasta que dichos Comités se constituyan, los Comités Éticos de Investigación Clínica que estén en funcionamiento en los centros que realicen investigación biomédica podrán asumir las competencias de aquéllos”. En concordancia con lo anterior, el último borrador del nuevo Real Decreto de Ensayos Clínicos, aún en trámite, establece una serie de requisitos adicionales que deberán cumplir aquellos CEI acreditados para la evaluación de ensayos clínicos con medicamentos y productos sanitarios, independientemente de que posean las competencias genéricas de los CEI.

La LIB crea además un segundo tipo de comités de ética con competencias en investigación biomédica, los **comités de ética externos de los biobancos, con un papel indudable en el terreno de los datos genéticos almacenados en régimen de biobanco.**

En el transcurso de una investigación biomédica que implique la participación de seres humanos, sus muestras o sus datos, el papel de los CEI se podría esquematizar tal como se muestra en la tabla 4.1.

En el ámbito concreto de los estudios genómicos, los CEI deberán evaluar particularmente los siguientes aspectos:

1. Que los participantes van a recibir información adecuada.
2. Que el consentimiento se otorga de manera libre por los participantes, con las garantías adicionales previstas en caso de ciertas poblaciones vulnerables o para los incapaces de consentir.

Tabla 4.1. Papel del comité de ética de investigación en el transcurso de una investigación biomédica

Análisis		Antes de comenzar la investigación		Después de que la investigación haya comenzado	
Fase de investigación		Planificación, diseño del proyecto		Revisión	
Funciones		Desarrollo		Fin de la investigación	
		Proponer información a los investigadores, cuando sea necesario		Evaluación ética* de la propuesta de investigación	
		Seguimiento del proyecto en aspectos éticos concretos; posible reevaluación		Revisar los informes de los investigadores	

*La evaluación ética incluye una comprobación de la adecuación a la normativa aplicable vigente y a los principios comúnmente aceptados en bioética, así como una ponderación del balance riesgo/beneficio de la investigación³³.

3. Que el consentimiento incluye posicionamiento del participante sobre el almacenamiento a largo plazo de sus datos, así como sobre la posible comunicación de hallazgos inesperados.
4. Que incluye posicionamiento del sujeto respecto a la salida de datos fuera del territorio nacional.
5. Que los procedimientos de disociación o, en su caso, anonimización, son adecuados, y que se garantiza en todo momento la confidencialidad.
6. Que el proyecto contempla los mecanismos para la adecuada transmisión de los resultados globales del estudio a los sujetos participantes.
7. Que se encuentra previsto un procedimiento para el caso en que se requiera comunicar a los sujetos sobre hallazgos inesperados, incluyendo un adecuado entorno de consejo genético.
8. Ante un hallazgo inesperado determinado, con posible repercusión sobre la salud del participante o sus familiares directos, la pertinencia de proceder a su comunicación, a la luz de lo establecido en el documento de consentimiento y el particular balance riesgo/beneficio de su conocimiento.
9. La pertinencia de los programas de cribado genético específicos, así como los supuestos en los que el consentimiento para acceder a un cribado genético podrá expresarse verbalmente.
10. Los casos excepcionales en que podrán tratarse muestras codificadas o identificadas con fines de investigación biomédica sin el consentimiento del sujeto fuente, cuando la obtención de dicho consentimiento represente un esfuerzo no razonable, o no sea posible porque el sujeto fuente hubiera fallecido o resultase imposible su localización.
11. Que se cumplen los restantes requisitos de la Ley 14/2007, de Investigación Biomédica. Por ejemplo, el CEI debe revisar también los acuerdos entre centros en caso de intercambio de muestras y datos, especialmente cuando tiene dimensión internacional. Se debe recordar que la LIB y el RD 1716/2011 son aplicables a la gestión de las muestras y los datos genéticos que se manejan en ensayos clínicos una vez que estos finalizan. Por lo tanto, los acuerdos en relación con su destino serán revisados a la luz de estas previsiones. Por otra parte, se puede sostener que no cabe exclusión para los ensayos clínicos de las reglas previstas en estas disposiciones para garantizar el respeto a los sujetos de la investigación³⁴.

Finalmente, es importante mencionar que el papel de los CEI más allá de la evaluación inicial de los proyectos, es decir, durante su desarrollo e incluso tras su finalización, debería potenciarse más. Se debería profundizar más en el ejercicio de su función de garantizar la continuidad de la relación entre los investigadores o los centros con los sujetos de la investigación, proponiendo y revisando vías de comunicación e identificando las responsabilidades que puedan subsistir. Algunos ejemplos en este sentido pueden ser la sugerencia sobre envío personalizado de resultados u otras de novedades, o la revisión de webs habilitadas para que determinada información esté disponible para los sujetos.

Papel de los CEI en la investigación con datos genéticos obtenidos a partir de bases de datos

El artículo 2 de la LIB señala que “La autorización y desarrollo de cualquier proyecto de investigación *sobre seres humanos o su material biológico* requerirá el previo y preceptivo informe favorable del Comité de Ética de la Investigación”. Este principio se concreta en el artículo 16 de la LIB para las investigaciones invasivas (que suponen un riesgo para la salud física o psíquica del sujeto –art. 3.t de la LIB–) y para las investigaciones con muestras en el artículo 62.

Es evidente, pues, que la evaluación previa de un CEI es imprescindible para iniciar cualquier investigación invasiva o con muestras biológicas, pero no lo es tanto cuando el investigador no requiere llevar a cabo una intervención en un ser humano ni analizar muestras. Este es el caso en que la investigación se vaya a llevar a cabo con datos genéticos almacenados resultantes de otras investigaciones (que sí fueron evaluadas) o de archivos clínicos.

Sin embargo, lo cierto es que la investigación con estos datos puede tener repercusiones para los derechos de sus titulares, y el artículo 5.1 exige que se garantice “la protección de la intimidad personal y el tratamiento confidencial de los datos personales que resulten de la actividad de investigación biomédica”.

Pues bien, parece que la función general atribuida a los CEIC en relación con la evaluación de la investigación debe abarcar también la supervisión de estas investigaciones, que únicamente requieren datos, para que se constate la garantía prevista en el artículo 5.1 de la LIB.

Por otra parte, el artículo 19 de la Declaración de Bioética de la UNESCO recomienda atribuir a los CEIC la función de supervisión de cualquier investigación, cuando señala que “los comités de ética deberían evaluar los proyectos relativos a seres humanos”. En el mismo sentido, la Declaración de Helsinki hace también referencia a la necesidad de evaluación de la investigación con datos identificables por parte de comités de ética (párrafo 32).

En otro orden de cosas, es importante señalar que es cada vez más frecuente que las publicaciones científicas exijan a los investigadores que sus estudios (cualquiera que haya sido la fuente de los datos) hayan sido revisados por un CEI.

No obstante, la evaluación del CEI para estos proyectos debería limitarse a los casos en que la investigación presentara un riesgo para los derechos de los sujetos: no se justificaría cuando se tratara de acceso a datos anónimos o anonimizados, como suele suceder cuando se accede a bases de datos de acceso público, y sí cuando se tratara de acceso a bases de datos genéticos

u otros archivos de acceso controlado. En el primer supuesto, de todas formas, el investigador podría optar por solicitar al CEI un informe sobre la no exigibilidad de evaluación por las características de su investigación.

En definitiva, la investigación con datos genéticos de sujetos identificables obtenidos de archivos o bases de datos debe ser sometida a evaluación de un CEI, pero es preciso que se establezca con claridad en la normativa en qué condiciones rige esta obligación.

RESUMEN DEL CAPÍTULO

El proceso de obtención, conservación, almacenamiento, utilización y cesión de los datos genéticos está regulado en nuestro ordenamiento jurídico de manera específica y detallada, puesto que es aplicable la normativa que regula estas operaciones cuando tienen por objeto muestras biológicas (LIB y RD 1716/2011). En lo no previsto en estas disposiciones se aplicará la legislación sobre protección de datos y sobre derechos de los pacientes y documentación clínica.

Independientemente del contexto en el que se obtengan, utilicen, almacenen o cedan los datos, la identificabilidad del sujeto es una circunstancia determinante para establecer qué requisitos deben cumplirse en su tratamiento. En última instancia, lo que se pretende es que el sujeto mantenga el control de la información que pueda ser relacionada con su identidad. La clave, por tanto, es la posibilidad de “trazabilidad” desde la información hasta un individuo concreto, cuando se considere que se puede establecer por medios razonables. A tal efecto deben tenerse en cuenta elementos como la finalidad del tratamiento, la manera en que el tratamiento está estructurado, el rédito que espera obtener el responsable del tratamiento, los intereses individuales en juego, así como el riesgo de que se produzcan disfunciones organizativas.

El test de la razonabilidad de la identificabilidad debe ser dinámico y tener en cuenta el contexto general y el conjunto de los datos disponibles por la comunidad científica. Las consecuencias de un almacenamiento masivo de información y del avance en los métodos para su análisis son factores que inciden en este sentido, como lo son también las medidas que se adopten para evitar la identificación: control en el acceso a las bases de datos, compromisos jurídicamente vinculantes entre los sujetos involucrados, controles externos, adopción de medidas de seguridad, información y consentimiento de los sujetos, transparencia en las políticas de gestión de las bases de datos, etc. Estas medidas deben ser proporcionales a las posibilidades de identificación según el contenido de las bases de datos (variantes génicas, mutaciones patológicas, mutaciones somáticas) y deben tener carácter dinámico, es decir, adaptables a la evolución de la técnica y al contexto en que se manejen. Según se apliquen estas medidas y condiciones, los datos genéticos archivados serán considerados anónimos y su tratamiento no estará sujeto a las mismas condiciones que las aplicables al de los datos de carácter personal.

Cuando se trata de bases de datos genéticos codificados (que son los disociados de manera reversible), la aplicación automática de todo el régimen previsto para el tratamiento de datos de carácter personal suscita algunas dudas que no han sido suficientemente discutidas. La presencia en cada caso particular de los elementos y las medidas aludidos en el párrafo anterior podría llevar a distintas soluciones (como de hecho ha ocurrido en el Código Tipo de Farmaindustria).

Tanto en el caso de datos anonimizados (cuando, por ser determinados datos genéticos, potencialmente podrían considerarse identificables aunque no por medios razonables) como en el caso de datos codificados, se debería valorar la oportunidad de impulsar un sistema de acreditación y registro de centros para su gestión. Este sistema sería más interesante si tuviera carácter internacional. Esta acreditación debería representar ventajas en la evaluación ética y jurídica de los proyectos que dichos centros llevaran a cabo. Ya existen algunas iniciativas en

el marco internacional en esta línea, como la constitución de la llamada Global Alliance for genomics and health.

Por lo que se refiere a la aplicación de los principios de protección de datos a las bases de datos genéticos de carácter personal, tiene una relevancia particular la determinación del periodo de su conservación. En el contexto clínico, podría estar justificada incluso la conservación indefinida. Sería conveniente que esta conservación respondiera a criterios concretos y definidos. En el contexto de la investigación, cuando la obtención de los datos se lleve a cabo en el marco de un proyecto concreto (se donaron o se cedieron las muestras para una investigación específica), la limitación del tiempo de almacenamiento al de la duración del proyecto puede no resultar adecuada. Ahora bien, el sujeto debe ser informado de las previsiones sobre la conservación y sus finalidades, y mantener la capacidad de decisión sobre ellas.

Es importante, asimismo, recordar que los derechos y obligaciones presentes en el proceso del tratamiento de los datos se distribuyen entre distintas categorías jurídicas diseñadas por la legislación sobre protección de datos. Es preciso encajar en estas categorías a los sujetos involucrados en la obtención, archivo y cesión de los datos genéticos para poder identificar qué derechos y obligaciones corresponden a cada uno. El titular de los datos es el sujeto fuente de la muestra, y sus familiares, en principio, no tienen la categoría de interesados o afectados: será él quien ostente el control sobre los datos y quien ejerza los derechos de acceso, rectificación o cancelación. El responsable del fichero o del tratamiento será la institución donde se gestionen los datos o el profesional que lo haga, según quién decida sobre la finalidad, contenido y uso del tratamiento, aunque no lo realizase materialmente. En el primer caso, los deberes como responsable corresponderán al titular del fichero del centro y el profesional tratará los datos como usuario autorizado. Cuando el responsable transfiera los datos a una persona o institución para que los gestione por cuenta de aquel, deberá establecerse una relación jurídica entre quien los trasfiere y quien los recibe, que será considerado encargado del tratamiento. Una de las principales modalidades de encargado del tratamiento es la prestación de servicios de *cloud computing*. El punto de partida para la contratación de estos servicios se encuentra en la exigencia de una diligencia cualificada por parte de quienes la llevan a cabo para informarse sobre la naturaleza de estos y sobre las garantías exigibles para cumplir con la normativa antes citada. Es particularmente importante recordar que en la prestación de estos servicios es habitual que intervengan terceras empresas subcontratadas, en cuyo caso el cliente que los contrata debe obtener la información sobre esta posibilidad y obtener garantías sobre las empresas subcontratistas. Además, las empresas que intervienen en la prestación de estos servicios (encargados del tratamiento o subcontratistas) pueden estar ubicadas en terceros países que no ofrezcan garantías adecuadas, por lo que es imprescindible obtener información sobre esta cuestión y sobre las que ofrecen para realizar transferencias internacionales de datos.

En todo este proceso se habrán de aplicar las medidas de seguridad contempladas en la LOPD y en el RLOPD 1720/2007, cuyo cumplimiento deberá garantizar el responsable del fichero. Este deber también corresponde al encargado del tratamiento. En particular es preciso insistir en la necesidad de que se lleve a cabo y se analicen los resultados del control de accesos y de la auditoría de seguridad. Esta medida de seguridad es especialmente relevante porque permite conocer el adecuado cumplimiento de las medidas de seguridad y porque, en la experiencia de la AEPD, se ha constatado que es una medida con un nivel reducido de cumplimiento.

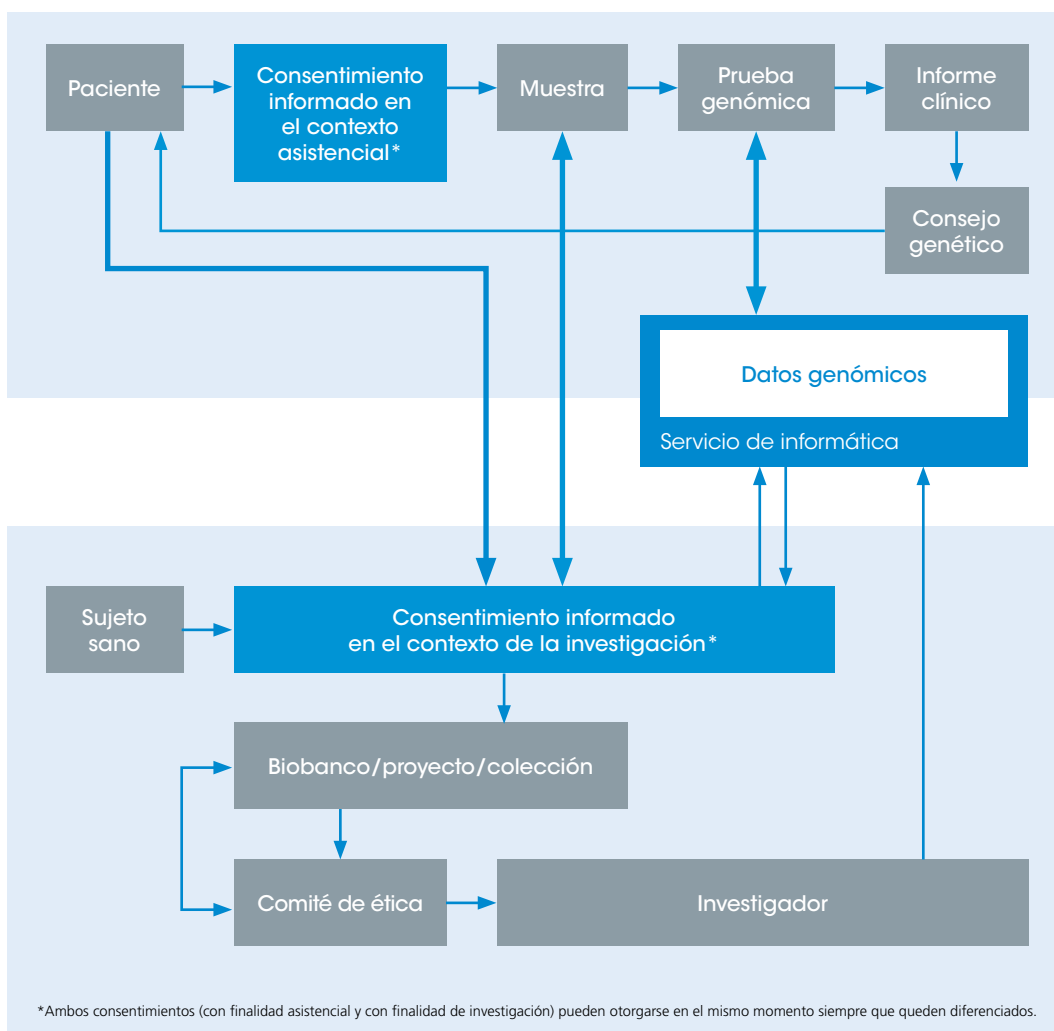
Junto con estas garantías, es imprescindible mencionar que las características específicas de la NGS, tales como la cantidad y la calidad de los datos y las oportunidades abiertas para futuros análisis y reanálisis de los datos, así como las nuevas posibilidades técnicas que se abren para su gestión, sugieren la especial relevancia de las políticas que se articulen para su tratamiento. La revisión de su puesta en práctica para cada escenario concreto por parte de expertos adquiere una gran relevancia.

El uso de la información genética en el contexto clínico puede generar conflictos de difícil gestión por parte del profesional y ser aconsejable recabar el consejo o recomendación del CEA, pero en la práctica esto es poco habitual. Se debería potenciar la función de los CEA, recogiendo en la normativa su papel en cuestiones éticas en el marco también de la gestión de los datos genéticos. Las instituciones y centros que trabajan con datos genéticos –en clínica y en investigación–, y que disponen de comités, deberían propiciar la relación entre ellos (CEI – CEI biobanco – CEA) a fin de actualizar periódicamente en qué se trabaja y cuáles son las cuestiones éticas implicadas.

La función de los CEI y de los comités de ética externos de los biobancos, en relación con la evaluación de la investigación, viene recogida expresamente en la LIB y en el RD 1716/2011. No obstante, si bien debe mantenerse que la investigación con datos genéticos de sujetos identificables obtenidos de archivos o bases de datos debe ser sometida a evaluación de un CEI, es preciso que se establezca con claridad en la normativa en qué condiciones rige esta obligación. Finalmente, es importante mencionar que el papel de los CEI más allá de la evaluación inicial de los proyectos, esto es, durante su desarrollo e incluso tras su finalización, debería potenciarse más. Se debería profundizar más en el ejercicio de su función de garantizar la continuidad de la relación entre los investigadores o los centros con los sujetos de la investigación, proponiendo y revisando vías de comunicación e identificando las responsabilidades que puedan subsistir. Algunos ejemplos en este sentido pueden ser la sugerencia sobre envío personalizado de resultados u otras de novedades, o la revisión de webs habilitadas para que determinada información esté disponible para los sujetos.



FLUJOS DE DATOS



En este capítulo se describirán las posibilidades de circulación de datos en el contexto clínico y de investigación, los sujetos involucrados en estos procesos, su calificación jurídica y los derechos y deberes que corresponden a cada uno.

Como punto de partida debe señalarse que la circulación de datos que suponga la revelación a una persona distinta del interesado, de los usuarios autorizados o del encargado del tratamiento es una operación de cesión (art. 3.i de la LOPD), que puede hacerse efectiva realizando una transferencia de información por parte del responsable del fichero a un tercero o permitiendo el acceso del tercero a la base de datos. En todo caso, la cesión de datos genómicos de carácter personal constituye un tratamiento de datos que debe estar sujeto a las condiciones establecidas en la LOPD y en la LIB.

IMPLICADOS EN LOS FLUJOS DE DATOS EN EL CONTEXTO DEL DIAGNÓSTICO GENÉTICO

El paciente como titular de los datos

INFORMACIÓN Y CONSENTIMIENTO PARA LA REALIZACIÓN DEL DIAGNÓSTICO GENÉTICO

Si bien cuando se trate de datos relativos a la salud en general no es preciso el consentimiento del interesado para su obtención si la finalidad del tratamiento es el diagnóstico clínico y se cumplen determinadas condiciones (arts. 7 y 8 de la LOPD), en el caso de datos genéticos la LIB dispone que el consentimiento por escrito es siempre necesario (art. 45.d de la LIB). Para que el consentimiento sea válido es preciso que se emita de forma libre, inequívoca, específica e informada, por una persona capaz o por su representante autorizado (art. 48.1 de la LIB). El consentimiento debe formalizarse por escrito, aunque en el caso de los cribados genéticos, el comité de ética de la investigación (CEI) puede avalar la posibilidad de que este sea expresado verbalmente (sorprendentemente, el artículo 48.3 de la LIB atribuye al CEI funciones para intervenir en un contexto asistencial). En todo caso, cuando el cribado incluya enfermedades no tratables o los beneficios sean escasos o inciertos, el consentimiento se obtendrá siempre por escrito (art. 48.1 de la LIB).

En cuanto al contenido de la información que ha de preceder al consentimiento, debe aclararse que si bien el artículo 47 se rubrica “Información previa a la realización de análisis genéticos con fines de investigación en el ámbito sanitario”, es aplicable a cualquier análisis en este ámbito, lo que se deduce de su remisión al artículo 48 en el que se establece la necesidad de consentimiento expreso para todo análisis genético en el ámbito sanitario (“Sin perjuicio de lo establecido en la legislación sobre protección de datos de carácter personal, antes de que el sujeto preste el consentimiento en los términos previstos en el artículo 48, deberá recibir la siguiente información por escrito”). El contenido de esta información es el siguiente:

1. Finalidad del análisis genético para el cual consiente.
2. Lugar de realización del análisis.
3. Destino de la muestra biológica al término del mismo, y previsión sobre el destino de la

muestra: destrucción, almacenamiento con fines asistenciales o almacenamiento con fines de investigación. En el último caso se solicitará el consentimiento correspondiente.

4. Personas que tendrán acceso a los resultados de los análisis cuando aquellos no vayan a ser sometidos a procedimientos de disociación o de anonimización.
5. Derecho a conocer los datos genéticos que se obtengan a partir del análisis de las muestras donadas.
6. Advertencia sobre la posibilidad de hallazgos inesperados y su posible trascendencia para el sujeto, así como sobre la facultad de este de tomar una posición en relación con recibir su comunicación.
7. Advertencia de la implicación que puede tener para sus familiares la información que se llegue a obtener y la conveniencia de que él mismo, en su caso, transmita dicha información a aquellos.
8. Compromiso de suministrar consejo genético, una vez obtenidos y evaluados los resultados del análisis.
9. Previsión de transferencia internacional de los datos genéticos. Si el sujeto fuente ha autorizado la cesión de datos a terceros sin mencionar explícitamente las transferencias internacionales de datos, dicho consentimiento permitirá realizarlas si se aportan garantías adecuadas.
10. Posibilidad de identificar a otras personas que ejerzan el derecho de acceso.

La identificación anticipada del lugar de realización del análisis y de las personas que tendrán acceso a los resultados (apartados 2 y 4) puede ofrecer importantes dificultades en los casos en que, pudiendo ser necesario acudir a otros centros o profesionales para cumplir adecuadamente la finalidad para la que se prestó el consentimiento, tales informaciones sean desconocidas. Si en el momento de obtenerse esta información resulta posible contactar con el sujeto fuente sin esfuerzos desproporcionados, no se plantearán mayores dificultades. De no ser así, cabría considerar la posibilidad de modular dichas garantías, informando previamente al sujeto fuente de que los lugares y profesionales que intervengan en los análisis lo harán vinculados a la finalidad habilitada por su consentimiento y poniendo a su disposición una opción para conocer dicha información (p. ej., un sitio web) en el que pueda concretarla y, en su caso, manifestar su negativa a consentirlo.

Si se trata de un análisis con fines de cribado genético, además de lo anterior, se deberá informar sobre lo previsto en el artículo 54.6 de la LIB:

1. Las características y objetivos que se persiguen con el cribado.
2. La naturaleza voluntaria de la participación.
3. La validez y fiabilidad de las pruebas de cribado y de las pruebas diagnósticas de segundo nivel.
4. La posibilidad de obtener falsos positivos y, en consecuencia, la necesidad de confirmar o descartar el diagnóstico.
5. Los periodos de tiempo que transcurrirán entre las distintas etapas del proceso del cribado.
6. Las posibilidades existentes de tratamiento y prevención de la enfermedad una vez diagnosticada.
7. Las incomodidades, riesgos y acontecimientos adversos que podrán derivarse del proceso diagnóstico, incluyendo los asociados a la toma de muestras y a las medidas terapéuticas o preventivas que ofrezca el programa.

A esto se debe añadir lo dispuesto en el artículo 5 de la LOPD en tanto sea pertinente en este contexto:

- “a) De la existencia de un fichero o tratamiento de datos de carácter personal, de la finalidad de la recogida de éstos y de los destinatarios de la información.
[...]
- d) De la posibilidad de ejercitar los derechos de acceso, rectificación, cancelación y oposición.
- e) De la identidad y dirección del responsable del tratamiento o, en su caso, de su representante”.

Finalmente, se debería valorar también la necesidad de ampliar la información sobre la posibilidad de incluir secuencias genéticas en bases de datos, su finalidad y las políticas para su gestión, incluso si se va a proceder a la anonimización. En cuanto a la realización de análisis clínicos de secuenciación de genomas completos, como se ha señalado, los puntos recogidos en la LIB parecen suficientes, en el sentido de que proporcionan la información adecuada para que el sujeto sea consciente de las características de estos análisis, siempre que se explique que se va a realizar esta técnica en particular³⁵.

DERECHO DE ACCESO

El derecho de acceso en el sentido del artículo 15 de la LOPD se refiere a cualquier dato genómico relativo al sujeto (datos crudos, información sobre variables genéticas, informes clínicos): “1. El interesado tendrá derecho a solicitar y obtener gratuitamente información de sus datos de carácter personal sometidos a tratamiento, el origen de dichos datos, así como las comunicaciones realizadas o que se prevén hacer de los mismos”. Está reconocido para los datos recogidos en la historia clínica (art. 18 de la Ley 41/2002) y para cualquier dato genético archivado en el ámbito biomédico en la LIB (arts. 49.1 y 32 del RD 1716/2011). Requiere de una solicitud por el interesado.

En relación con el acceso a los datos genéticos, se podría contemplar la necesidad de intervención de un profesional para asegurar su comprensión (consejo genético), en lugar de una comunicación directa, o al menos advertir de su trascendencia.

El interesado debe ser informado de cómo ejercer este derecho y se deberá referir también a las cesiones, sea que hubiera consentido expresamente a cada una de ellas o que este consentimiento específico no se hubiera requerido por cualquier razón (se encontraba englobado en el consentimiento inicial, el comité de ética entendió que no era necesario, se cedieron datos no identificativos...).

La obligación de satisfacer este derecho corresponde al responsable del fichero.

Se han planteado algunos argumentos para establecer ciertas limitaciones a este derecho, como, entre otros, cuando los datos no tengan relevancia clínica conocida, cuando el sujeto pueda resultar perjudicado según las condiciones de dicho acceso o cuando la entrega de la información suponga un esfuerzo desproporcionado. No parece que sean sostenibles los dos primeros argumentos: aunque los datos no tengan relevancia clínica conocida, son datos de

carácter personal objeto de este derecho; en cuanto a la excepción terapéutica (art. 5.4 de la Ley 41/2002), cede a favor de la solicitud expresa del sujeto. En relación con el tercero (la entrega de la información supone un esfuerzo desproporcionado), debe tenerse en cuenta que, aunque el artículo 28.1 del RLOPD reconoce al interesado el derecho a optar por la modalidad de acceso, si el responsable ofrece un procedimiento para hacer efectivo el acceso con efectos y seguridad adecuados y el interesado exigiese otro que implique un coste desproporcionado, los gastos derivados de su elección serán por cuenta del afectado (art. 58.3 del RLOPD). En todo caso, debe recordarse que cuando los datos se almacenan por un prestador de servicios (encargado del tratamiento), deben articularse medidas que permitan el ejercicio del derecho por el responsable (p.ej., si se contratan servicios de almacenamiento en la nube).

REVOCACIÓN DEL CONSENTIMIENTO Y DERECHO DE CANCELACIÓN

La LIB recoge diversas previsiones sobre el consentimiento del sujeto fuente y su revocación. Sin embargo, en lo no previsto en dicha norma, la regulación de la prestación del consentimiento y su revocación debe completarse con lo establecido en la LOPD.

En este sentido, la LOPD distingue entre la revocación del consentimiento y el ejercicio de los derechos de cancelación y oposición.

La revocación del consentimiento, cuando este es necesario para legitimar el tratamiento de los datos, impide que el mismo se siga realizando, incluso para proceder a su anonimización, ya que esta implica un nuevo y posterior uso de la información personal.

Por su parte, el ejercicio del derecho de cancelación supone, como regla general, el bloqueo de los datos, por lo que debe conservarse exclusivamente a disposición de las administraciones públicas, jueces o tribunales, para la atención de las posibles responsabilidades nacidas del tratamiento de los datos y solo durante el plazo de prescripción de estas (arts. 31 y 5.1.b del RLOPD). Transcurrido dicho plazo, debe procederse a la supresión (art. 16.3 de la LOPD). Sería recomendable informar a los interesados sobre el significado del “bloqueo” y su diferencia con la “cancelación” y la “supresión”. No obstante, debe tenerse en cuenta que el derecho de cancelación puede ser denegado motivadamente en distintos supuestos.

En el ámbito de la información clínica, el consentimiento, su revocación y el derecho de cancelación presentan singularidades derivadas de la Ley 41/2002.

Así, la incorporación y el tratamiento de los datos de la historia clínica no precisan del consentimiento del paciente al concurrir habilitación y exigencias legales sobre su cumplimentación, finalidades, accesos y conservación (arts. 15 y ss Ley 41/2002). Por tanto, tampoco cabe la revocación del consentimiento en la medida en que deba cumplirse con dichas exigencias legales. Asimismo, el derecho a la cancelación de los datos contenidos en la historia clínica queda limitado por las obligaciones de conservación de esta documentación en la Ley 41/2002 y en las distintas normas de las comunidades autónomas que pueden contemplar periodos distintos de conservación, en función del tipo de información contenida en la historia clínica. El plazo de conservación de 5 años establecido en la Ley 41/2002 es el mínimo exigible, pero puede ampliarse si es necesario “para la debida asistencia al paciente durante el tiempo adecuado en cada caso” (art. 17.1 Ley 41/2002).

Para los datos genéticos existe una previsión específica en la LIB que señala la obligación de conservación durante un mínimo también de 5 años (art. 52). Igual que ocurre con los datos de salud en general archivados en la historia, la conservación de los genéticos está supeditada a las previsiones de las normas autonómicas específicas o generales para la documentación clínica.

DERECHO A LA INFORMACIÓN SOBRE LOS RESULTADOS Y DERECHO A NO SABER

Cercano al derecho de acceso a los datos, pero de naturaleza distinta, se reconoce al sujeto el derecho a ser informado sobre los resultados de los análisis genéticos que sean relevantes para su salud o la de sus familiares. Este derecho emana del derecho a la protección de la salud y hace surgir el correspondiente deber para los profesionales y los centros.

El deber de informar se refiere a datos científicamente validados y que puedan tener relevancia para la salud del sujeto (entendida en sentido amplio) o la de sus familiares (sin perjuicio del derecho de acceso a cualquier resultado, que requiere una solicitud específica posterior por parte del sujeto).

En relación con la valoración de la relevancia de la información y la posible limitación de su comunicación, cabe mencionar la previsión del artículo 5.4 de la Ley 41/2002, que prevé la posibilidad de la “excepción terapéutica” en materia de información: “el derecho a la información sanitaria de los pacientes puede limitarse por la existencia acreditada de un estado de necesidad terapéutica. Se entenderá por necesidad terapéutica la facultad del médico para actuar profesionalmente sin informar antes al paciente, cuando por razones objetivas el conocimiento de su propia situación pueda perjudicar su salud de manera grave. Llegado este caso, el médico dejará constancia razonada de las circunstancias en la historia clínica y comunicará su decisión a las personas vinculadas al paciente por razones familiares o de hecho”. Ahora bien, no parece fácil que pueda haber un “estado de necesidad terapéutica” (razones objetivas que avalen que conocer la información puede perjudicar gravemente su salud), que justifique ocultar información genética al paciente. La previsión del legislador parece basarse más bien en una situación límite, excepcional (p. ej., en paciente con riesgo de suicidio) y siempre compartida con el entorno del paciente y documentada en la historia clínica, sin perjuicio de un deber de información posterior, una vez superado el estado de necesidad.

En cualquier caso, si así se valorase y se aplicase la excepción del derecho a la información genética basándose en este precepto, debería ser siempre con carácter extraordinario y no generalizable. En la cumplimentación del deber de información en cuanto a los datos genéticos, la relevancia para la salud debe ser valorada en términos de gravedad, posibilidades de tratamiento presente o futuro, repercusión en la toma de decisiones reproductivas o de otro tipo, etc.

El objeto de este derecho es tanto la información que se busca (lo que no plantea problema puesto que es la finalidad de la prueba) como aquella que no se busca pero se sabe que se va a encontrar (hallazgos secundarios), e incluso la que se encuentra inesperadamente (hallazgos inesperados).

Los hallazgos inesperados son información sobrevenida, no buscada, ni esperada, dentro del objetivo inicial del análisis genético, pero que puede aportar datos relevantes para el sujeto fuente.

La mayor dificultad en la gestión de los hallazgos inesperados, que eventualmente se presenta en diferentes pruebas médicas, tiene un especial interés en el contexto de los análisis genéticos. Es una cuestión que preocupa y a la que cada vez se presta más atención³⁶.

Como se dijo, el ordenamiento jurídico reconoce el derecho del sujeto a controlar la información que le concierne, así como a tomar decisiones libremente en el contexto sanitario. Por esta razón, se debe respetar su decisión en relación con la comunicación o no de esta información.

Para garantizar el respeto a la autonomía del sujeto y por lo tanto la capacidad de elección entre “saber” o “no saber”, se debe anticipar al sujeto fuente que de ese análisis genético realizado para verificar cierto diagnóstico o alteración pueden surgir nuevos datos no esperados y consultarle sobre su voluntad de conocerlos. De este modo, se cubre un doble objetivo ético-legal: respetar su derecho a “no saber” y prepararle para una futura información que pudiese ser relevante. No hacerlo supone someter posteriormente al sujeto, si se constata el “hallazgo”, a mayor estrés y ansiedad por lo desconocido. Por otro lado, hay que tener en cuenta que la información surgida del análisis genético, buscada o hallada inesperadamente, puede ser trascendente no solo para el propio sujeto fuente, sino para terceros familiares. En tal caso, hay que contrastar con él la necesidad de compartir esta información con los posibles “afectados” y la posibilidad de tener que comunicarla a estos en caso de su negativa a saber (v. pág. 83).

Es importante que todo este proceso quede documentado y que se identifique a un familiar o allegado del sujeto para que pueda ser contactado en caso de hallazgos que pudiesen afectar a terceros y aquel haya ejercido su derecho a la ignorancia.

En el escenario que acabamos de describir se puede presentar un dilema cuando las nuevas evidencias científicas plantean la posibilidad de reconsiderar un diagnóstico a partir de muestras o secuencias conservadas. El deber del médico con su paciente supone adaptar el diagnóstico y el tratamiento al estado de la ciencia y por tanto abarca la actualización del diagnóstico inicial. Si para actualizar el diagnóstico fueran precisas nuevas pruebas genéticas sobre cuyas implicaciones no se hubiera informado al paciente, debería contactarse con él con carácter previo para solicitar su consentimiento. Los comités de ética asistencial pueden desempeñar un papel importante a la hora de tomar decisiones en este sentido (v. pág. 83 en relación con el deber de revisar los diagnósticos).

EL PACIENTE MENOR

La participación de los menores en las decisiones relativas a diagnósticos y tratamientos es un tema debatido en el ámbito jurídico. La tendencia evoluciona a favor de una implicación de los menores de acuerdo a su capacidad, y así, como principio general, se puede afirmar que “el paciente participará en la medida de lo posible en la toma de decisiones a lo largo del proceso sanitario” (art. 9.5 de la Ley 41/2002 básica reguladora de la autonomía del paciente y de derechos y obligaciones en materia de información y documentación clínica).

El paciente menor es quien debe ser informado y consentir en relación con las actuaciones en el ámbito sanitario cuando su grado de madurez lo permita. Se entiende que, cuando se han cumplido dieciséis años, se reúnen ya las condiciones de madurez necesarias. Por debajo

de esa edad, el médico implicará al menor en la información y en la toma de decisiones en función de la apreciación de su madurez. No obstante, cuando se trate de una actuación de grave riesgo y a criterio del facultativo, los padres serán informados y su opinión será tenida en cuenta para la toma de la decisión correspondiente.

Estos criterios generales deben ser aplicados a los análisis genéticos diagnósticos; es decir, los menores deben estar informados y consentir por sí mismos, según su capacidad, que en el contexto clínico se entiende adquirida a los dieciséis años por regla general. Sin embargo, la LIB prevé que cuando se lleve a cabo un estudio genético familiar, los datos que resulten de los análisis de los menores sean comunicados a sus tutores o representantes legales (art. 51.2 de la LIB). Esto debería ser tenido en cuenta a la hora de informar a los menores, y debería advertírseles de la futura comunicación de los datos.

En cuanto a la indicación de los análisis, la Orden SSI/2065/2014, de 31 de octubre, por la que se modifican los anexos I, II y III del Real Decreto 1030/2006, de 15 de septiembre, por el que se establece la cartera de servicios comunes del Sistema Nacional de Salud y el procedimiento para su actualización, señala que: "En el caso de menores, el estudio genético presintomático en enfermedades que aparecen en la edad adulta deberá diferirse hasta que la persona tenga la madurez y competencia necesaria para comprender la naturaleza e implicaciones de su decisión, salvo que existan medidas preventivas eficaces aplicables en la infancia. En enfermedades que aparecen en la infancia y que pueden ser prevenidas o tratadas adecuadamente se deberá realizar lo más cercano posible a la fecha en la que se deben iniciar dichas medidas preventivas y/o terapéuticas [...] El estudio genético de portadores "no se realizará hasta que el menor alcance la madurez y competencia necesarias para comprender la naturaleza de su decisión y sus implicaciones y sea, a su vez, capaz de dar su consentimiento".

Debido al riesgo de aparición de hallazgos inesperados, es muy recomendable restringir el uso de NGS (*Next Generation Sequencing*) en menores con finalidades diagnósticas o de investigación no predictivas y focalizar en lo posible el análisis de variantes o genes relacionados con la patología usando los filtros adecuados.

En cuanto al almacenamiento de los datos genéticos del menor es obligado en función de lo previsto en la Ley 41/2002 para el archivo de la documentación clínica. Sus representantes legales ejercerán los derechos correspondientes hasta que el paciente alcance la mayoría de edad o adquiera la madurez suficiente. Ahora bien, su utilización con otros fines debe quedar justificado específicamente y sustentarse en garantías adicionales, como se verá en el apartado "Implicados en los flujos de datos en el contexto de la investigación (v. pág. 85).

DIAGNÓSTICO PRENATAL

Uno de los escenarios posibles en diagnóstico genético, con afectación posible a terceros, es el que puede darse en diagnóstico prenatal, e incluso con carácter previo a este, en el contexto de consejo genético reproductivo en parejas con antecedentes familiares de enfermedades genéticas hereditarias.

En diagnóstico prenatal cobra especial importancia el tiempo y modo en que debe darse a los progenitores la información genética referente al feto y la dificultad pronóstica que a me-

nudo acompaña este tipo de diagnósticos, especialmente en patologías graves, que pueden conllevar afectaciones graves que acortan el pronóstico vital del recién nacido, caso de seguir adelante con el embarazo. Precisamente, de esa información suele derivarse una decisión tan difícil como es la de seguir o no con el embarazo, siempre debiendo respetar el marco jurídico vigente. La gestión adecuada de la información y el registro de este diagnóstico genético plantea doble dificultad por cuanto se consignará inicialmente en la historia clínica de la embarazada, pero lógicamente, debería también recogerse posteriormente en la historia clínica del recién nacido si el embarazo sigue adelante hasta el nacimiento. Si es así, estamos replicando la información sensible en dos registros distintos, con las consiguientes cautelas que deberían tenerse en la preservación de su confidencialidad.

En este punto sería bueno verificar si realmente de esta información genética que afecta al recién nacido queda constancia en su historia clínica, pues en otro contexto como sería la reproducción asistida, donde *a priori* no hay patología, el antecedente de su procedimiento de gestación muy a menudo no queda recogido y consta únicamente en el historial ginecológico de la madre gestante.

Tanto el registro de los antecedentes de reproducción asistida con material de donante como la existencia de ciertos datos genéticos que afecten al feto, relevantes para el futuro, deberían constar claramente identificados en su historia clínica, planteándose entonces otra cuestión no menos compleja como sería el acceso a esos datos por parte del afectado, es decir, el nacido. En cuanto al origen genético, en el caso de haber nacido de donante de gametos, es un tema sometido al criterio de confidencialidad del donante y por tanto al anonimato, según tiene establecida la propia legislación en materia de reproducción asistida en España. Sin embargo, se prevé la excepción de poder tener acceso a esos datos del donante en caso de aparecer enfermedad que pueda estar vinculada a ese origen genético (si bien no por parte del descendiente, sino de manera que se pueda solventar el problema de salud).

Respecto a los datos genéticos predeterminados ya en fase fetal, igualmente deberían trasladarse a la historia clínica del recién nacido, que debería tener acceso a ellos al alcanzar su mayoría de edad, del mismo modo que establecemos ese criterio en el apartado anterior, salvo que se trate de información relevante precozmente, de la que pueda derivarse un tratamiento o beneficio para el menor en su etapa de infancia o adolescencia.

En este contexto de patología genética determinada en fase prenatal, el criterio de dar o no acceso al menor, compartir o no con él esa información, debería estar siempre regido por el criterio de su mejor interés y potencial beneficio en esa información, desde la perspectiva diagnóstica y terapéutica. Es aconsejable un criterio de prudencia en esta cuestión, dada la especial dificultad en su comprensión y asimilación, según gravedad de los datos a facilitar.

Así, los profesionales que tienen bajo su responsabilidad realizar ese proceso de información a partir de los resultados de las técnicas de diagnóstico prenatal, orientadas a la identificación genética, tras la confirmación del diagnóstico, deben contar con una formación específica a fin de acompañar el proceso, psicológicamente duro para los progenitores, y a su vez mantenerse objetivamente neutros y no pretender influir en la decisión de la pareja. Debemos tener en cuenta que más allá de la decisión frente a ese embarazo en curso, la información que se proporcione puede afectar a decisiones futuras respecto a la descendencia de la pareja o incluso afectar a otros hijos ya nacidos que pueden verse afectados de patología aún no identificada.

Los análisis genéticos prenatales se realizarán cuando exista un alto riesgo de sufrir una determinada enfermedad o trastorno genético relacionado con la salud del feto siempre que:

- “i) El feto tiene alto riesgo de padecer una enfermedad o trastorno genético graves, o sus progenitores pertenecen a una familia en la que se ha descrito la presencia de una enfermedad o trastorno genético graves; ii) La enfermedad o trastorno genético tiene una alteración genética conocida y que puede ser identificada mediante el correspondiente análisis genético; iii) El análisis genético debe contribuir al manejo clínico de la gestación o del recién nacido o a la toma de decisiones reproductivas” (Orden SSI/2065/2014, citada).

Cuando la consulta genética es con carácter previo a un embarazo deseado, por razón de antecedentes familiares, a fin de intentar asegurar que este será sin riesgo de enfermedad para el que va a nacer, estamos ante el llamado “consejo genético reproductivo”, que requiere igualmente de profesionales formados específicamente para ello, para orientar la toma de decisiones ante las distintas alternativas. Esto debería ser realizado por profesionales formados según las recomendaciones de los organismos oficiales pertinentes (p. ej., el Ministerio de Sanidad, etc.). El elemento diferencial básico es que no tenemos todavía embarazo en curso y por tanto no deben tomarse decisiones sobre una vida en formación. Así, el rol que deben tener los profesionales es muy importante. La información obtenida quedará también registrada en la historia clínica de la(s) persona(s) consultantes (sean estos afectados o portadores sanos potenciales), si bien en caso de generarse un embarazo, será recogido también en el historial ginecológico de la embarazada.

Por tanto, sea como derivada de una prueba de diagnóstico prenatal o sea información generada a partir del consejo genético reproductivo, los datos obtenidos se recogerán probablemente en más de un registro, lo que requiere mayor cautela aún en su manejo para preservar su confidencialidad, debiendo tener especial cuidado con los datos que afectarán al futuro hijo, desde su mismo nacimiento y por el resto de su vida. Se debe recordar que el artículo 53 de la LIB establece que los resultados de los análisis genéticos realizados en material embrionario o fetal estarán sometidos a los principios de protección de datos y de confidencialidad establecidos en esta Ley.

EL PACIENTE FALLECIDO

En el ámbito sanitario es posible tanto acceder a historias clínicas de pacientes fallecidos (art. 18.4 de la Ley 41/2002) como obtener y/o analizar sus muestras, siempre que pueda resultar de interés para la protección de la salud de terceros (art. 48.2 de la LIB). El acceso en estos casos queda justificado por una finalidad legítima prevista en la ley, que articula sistemas para proteger la información contenida en las historias clínicas también tras el fallecimiento de los pacientes.

Por otra parte, estos mismos artículos recogen una excepción a la posibilidad de acceso a la historia en el caso de que el fallecido lo hubiera prohibido expresamente en vida y así se acredite. Ante esta prohibición se debería llevar a cabo una ponderación de intereses, y no se debe descartar una actuación en contra si el perjuicio que pudiera causar dicha prohibición tuviera relevancia para la salud de los familiares. También en esta cuestión, en caso de conflicto, puede ser de utilidad la intervención del comité de ética, asistencial o de investigación según el contexto.

Los familiares del paciente

Como se dijo, el titular del derecho a la autodeterminación en relación con los datos es el paciente del que proviene la muestra que se ha analizado. A pesar del carácter heredable de la información genética, solo a él corresponde el ejercicio de las facultades propias del titular de los datos. Sus familiares, en principio, no tienen la categoría de interesados o afectados, aunque pudieran compartir con este la información genética (cosa que no se sabe con certeza hasta que se les realicen los correspondientes análisis). El archivo de la documentación genética de los miembros de una familia debe llevarse a cabo por separado en las historias correspondientes (art. 51.2 de la LIB).

El artículo 5.2 de la LIB dispone que “En el supuesto de que los datos obtenidos del sujeto fuente pudieran revelar información de carácter personal de sus familiares, la cesión a terceros requerirá el consentimiento expreso y escrito de todos los interesados”. En este caso, la involucración de los familiares deberá estar justificada por pasar a considerarse como interesados, esto es, como sujetos identificables de una información que se sabe que les corresponde.

Por otra parte, este hecho se tiene en cuenta también en la gestión de la información, como interés que pesa en la resolución de ciertos conflictos que pudieran surgir, pero sin que se justifique el reconocimiento de un correspondiente derecho subjetivo sobre la información. Por ejemplo, cuando el titular opta por no conocer los resultados de un análisis pero estos pueden ser relevantes para la salud de sus familiares, se articulan mecanismos para hacer llegar a estos la información. Además, como se dijo, el derecho a la protección de la salud y a la integridad física puede justificar excepcionalmente la ruptura del deber de secreto en una situación de estado de necesidad. El papel de los comités de ética asistencial puede ser muy relevante en este sentido: deberían intervenir para ponderar los intereses en el caso concreto, incluida la posibilidad de que los familiares prefirieran mantenerse en la ignorancia, y para solucionar aspectos concretos que se plantean en esta situación, como por ejemplo quién y cómo se contacta con los familiares.

El clínico que solicita el análisis y la institución sanitaria

El clínico que solicita el análisis y que luego recibe el informe accede a esta información legitimado por el ejercicio de su obligación asistencial (art. 50.1 de la LIB).

El archivo de esta documentación es responsabilidad del centro sanitario, tal como prevé la Ley 41/2002. En el supuesto de que contrate la gestión de la documentación clínica con una entidad externa se deberá formalizar el contrato correspondiente por el que dicha entidad pasará a ser un “encargado del tratamiento”.

En todo caso, existe una obligación de mantenimiento de los registros de los datos generados, incluso los registros de datos crudos: el conocimiento de la genómica y la aplicación clínica de los hallazgos está en constante evolución y es posible que los avances en esta área permitan una reevaluación de los resultados para mejorar el diagnóstico y/o tratamiento del paciente. Esta posibilidad de reanálisis solo podrá llevarse a cabo si se almacenan estos datos.

Esta nueva evaluación, si es que deviene clínicamente indicada, constituirá un deber del médico responsable durante el proceso asistencial, que estará legitimado por tanto para acceder a los re-

gistros correspondientes por razones asistenciales. La importancia de la información así obtenida para el diagnóstico y/o el tratamiento del paciente puede modificar el alcance de la obligación de contactar e informar por parte del médico, incluso con posterioridad a que el paciente haya sido dado de alta y/o haya cambiado el médico responsable. Por ello, los nuevos tratamientos de los datos de salud, y en concreto la información genética, exigen una reflexión más detallada sobre las obligaciones, los derechos y las responsabilidades de los profesionales asistenciales y los usuarios en relación con la gestión y el acceso a dicha información, que hasta la fecha se ha ocupado principalmente de los usos actuales o conocidos en el momento de la obtención de la información, pero no de usos potenciales o desconocidos en dicho momento inicial.

El personal del laboratorio que realiza la prueba y la institución que alberga el laboratorio. Posibilidad de solicitud de pruebas al extranjero

Es frecuente que los análisis genéticos se lleven a cabo en centros distintos de aquel al que pertenece el clínico responsable del paciente.

El centro que dispone de la tecnología, equipamiento y personal adecuado para llevar a cabo este análisis presta un servicio que deberá formalizarse en un contrato en el que se describan las condiciones correspondientes, entre las que deben constar las relativas a los datos que se obtienen: a efectos de la LOPD, la circulación de muestras y datos no se constituye en una cesión, sino en un tratamiento de datos por cuenta de terceros que pasa a ser un “encargado del tratamiento” (v. pág. 59). No es preciso, por tanto, el consentimiento del paciente para que se lleve a cabo esta operación. En estos acuerdos debería figurar:

- El objeto del análisis.
- La obligación de confidencialidad.
- La obligación de no realizar análisis ni utilizar los datos con fines distintos a los acordados.
- La obligación de tratar los datos con medidas de seguridad proporcionadas.
- La obligación de mantener los datos únicamente durante el periodo correspondiente al cumplimiento de las obligaciones legales.
- La posibilidad de almacenamiento de datos en el centro, si así se acuerda, por cuenta del centro hospitalario y el modo en que se garantizará el derecho de acceso y cancelación por parte del titular de los datos.

Si bien esta operación no se considera una cesión a efectos de la LOPD, si implica una transferencia internacional a terceros países sin garantías adecuadas, deberá ser autorizada por la Agencia Española de Protección de Datos y notificarse al Registro General de Protección de Datos.

El profesional que lleva a cabo la interpretación clínica de los resultados y otros profesionales de la institución

Como se dijo, los profesionales sanitarios acceden a la información clínica de sus pacientes legitimados por el ejercicio de su obligación asistencial.

Puede ocurrir que el profesional que lleva a cabo la interpretación clínica de los resultados no tenga una relación asistencial con el paciente y lo mismo ocurre con otros profesionales de la

institución donde se almacena la información. Por otra parte, puede ser pertinente distinguir entre perfiles profesionales específicos para diferenciar niveles de acceso en relación con la información genética. La informatización de esta documentación permitiría diferenciar y discriminar estos niveles.

Sería adecuado que la información se administrara por una persona autorizada (preferiblemente un bioinformático) que fuera capaz de otorgar distintos permisos a diferentes usuarios. Por ejemplo, un especialista en genética podría tener permiso para solicitar o ejecutar un tipo de análisis determinado, que puede ser el análisis de la línea germinal de un paciente, el análisis del tumor para identificar mutaciones somáticas, el análisis de una familia con el objeto de identificar la causa de una enfermedad hereditaria,... mientras que otros profesionales podrían tener permiso para visualizar los resultados relacionados con su especialidad o pacientes, o solicitar un análisis concreto para que sea llevado a cabo por el servicio correspondiente, pero no para ejecutar análisis por sí mismos, a menos que tengan acreditado un conocimiento suficiente del sistema y del tipo de datos, etc.

IMPLICADOS EN LOS FLUJOS DE DATOS EN EL CONTEXTO DE LA INVESTIGACIÓN

El sujeto de la investigación como titular de los datos

Como en el contexto clínico, en el de la investigación biomédica los derechos sobre los datos corresponden a su titular, que es el sujeto fuente de la muestra biológica.

INFORMACIÓN Y CONSENTIMIENTO PARA LA OBTENCIÓN Y UTILIZACIÓN DE ANÁLISIS GENÉTICOS CON FINES DE INVESTIGACIÓN

Por lo que se refiere a la obtención de datos genéticos con fines de investigación a partir del análisis de muestras biológicas, la LIB recoge la exigencia de consentimiento expreso, incluso si los datos se van a anonimizar (art. 58.2): el consentimiento no solo es expresión de la voluntad en relación con los datos de carácter personal, sino también en relación con la utilización de las partes del cuerpo separadas del mismo.

Por otra parte, la LIB señala que el consentimiento para el tratamiento de datos es siempre necesario salvo que estos hayan sido *previamente* anonimizados (art. 50.2). Por otra parte, se exige el consentimiento del sujeto fuente de la muestra biológica para que esta sea analizada con fines de investigación, incluso si se va a proceder a su anonimización (art. 58.2 de la LIB). Parece, pues, que se prevé que el sujeto fuente decida sobre la futura anonimización de sus datos genéticos.

La información se deberá referir a los siguientes puntos (art. 59 de la LIB):

- Finalidad y beneficios esperados del proyecto, línea de investigación o biobanco, e indicación de que la muestra y los datos solo podrán ser utilizados en el ámbito de esa finalidad. Pueden preverse no varias finalidades y que el consentimiento se refiera a todas ellas, si

es que están suficientemente descritas (p. ej., explicar que cuando acabe el proyecto, la muestra y los datos se incorporarán a una colección o a un biobanco). Pueden también preverse restricciones sobre el uso de la muestra o los datos, que tendrán más sentido cuando su destino sea un biobanco, puesto que el consentimiento se otorga en términos más amplios.

- Beneficios esperados del proyecto, colección o biobanco.
- Posibles inconvenientes vinculados con la donación y obtención de la muestra, incluida la posibilidad de ser contactado con posterioridad con el fin de recabar nuevos datos u obtener otras muestras.
- Identidad del responsable de la investigación, de la colección o descripción del biobanco.
- Derecho de revocación del consentimiento, forma de ejercerlo y sus efectos, incluida la posibilidad de la destrucción o de la anonimización de la muestra o de los datos y de que tales efectos no se extenderán a los datos resultantes de las investigaciones que ya se hayan llevado a cabo.
- Lugar de realización del análisis y destino de la muestra y los datos al término de la investigación: disociación, destrucción u otras investigaciones, y que en su caso comportará a su vez el cumplimiento de los requerimientos previstos en la Ley. En el caso de que estos extremos no se conozcan en el momento, se establecerá el compromiso de informar sobre ello en cuanto se conozcan. En el caso de que el destino de la donación sea un biobanco se deberá advertir que en caso de cierre, la información sobre el destino de las muestras estará a su disposición en el Registro Nacional de Biobancos para Investigación Biomédica con el fin de que pueda manifestar su conformidad o disconformidad con el destino previsto, todo ello sin perjuicio de la información que deba recibir por escrito el sujeto fuente antes de otorgar su consentimiento para la obtención y utilización de la muestra. En relación con la identificación de los centros, es un requisito que se debe entender efectivo cuando se establezcan garantías funcionalmente semejantes como que la identificación estará a disposición del sujeto si la solicita.
- Derecho a conocer los datos genéticos que se obtengan a partir del análisis de las muestras donadas y forma de ejercerlo. Indicación de que el biobanco y la persona responsable de la colección o proyecto de investigación tendrán a disposición del donante toda la información sobre los proyectos de investigación en los que se utilice la muestra y de que el comité de ética externo del biobanco o el CEI que evaluó el proyecto de investigación decidirán en qué casos será imprescindible que se envíe la información de manera individualizada.
- Garantía de confidencialidad de la información obtenida, indicando la identidad de las personas que tendrán acceso a los datos de carácter personal del sujeto fuente. Este requisito se debe entender efectivo cuando se establezcan garantías funcionalmente semejantes, como que la identidad de estos sujetos estará a disposición del sujeto si la solicita.
- Advertencia sobre la posibilidad de que se obtenga información relativa a su salud derivada de los análisis genéticos que se realicen sobre su muestra biológica, así como sobre su facultad de tomar una posición en relación con su comunicación.
- Advertencia de la implicación de la información que se pudiera obtener para sus familiares y la conveniencia de que él mismo, en su caso, transmita dicha información a aquellos.
- Indicación de la posibilidad de ponerse en contacto con él/ella, para lo que podrá solicitarle información sobre el modo de hacerlo.
- Renuncia a cualquier derecho de naturaleza económica, patrimonial o potestativa sobre los resultados o potenciales beneficios que puedan derivarse de manera directa o indirecta de las investigaciones que se lleven a cabo con la muestra que cede para investigación. No

obstante, y sin perjuicio de lo establecido en el artículo 7 de la Ley 14/2007, de 3 de julio, podrá fijarse una compensación económica por las molestias físicas, los gastos y otros inconvenientes que puedan derivarse de la toma de la muestra.

- En el caso de almacenamiento de muestras o datos de menores de edad, garantía de acceso a la información indicada en el artículo 32 del Real Decreto sobre la muestra por el sujeto fuente cuando este alcance la mayoría de edad.

La información que debe transmitirse al sujeto en caso de que las muestras o los datos se vayan a anonimizar será únicamente la relativa al eventual riesgo físico de la obtención, a cuestiones generales sobre el propósito y circunstancias de la investigación y a las consecuencias de la anonimización, en particular, sobre la imposibilidad de ejercer los derechos ARCO.

El detalle con que describen la LIB y el RD los puntos sobre los que se debe informar al donante no debe conducir a olvidar otros aspectos y a reducir el proceso de información y el acto de consentimiento a meros formalismos requeridos para que el comité de ética evalúe positivamente la investigación. No ha de olvidarse que el fundamento de la información es asegurar que el consentimiento se otorga de forma válida, una vez que el sujeto es consciente de “para qué” está consintiendo y de la trascendencia de la decisión. Es imprescindible que la información se otorgue en términos adecuados, es decir comprensibles, oralmente (aunque conste por escrito) y dando la oportunidad de hacer preguntas. No se debe desvirtuar esta finalidad entrando en detalles o explicaciones técnicas que dificulten la comprensión o elaborando formularios tan extensos que desincentiven su lectura.

Como ya se ha señalado (v. pág. 74), se debería valorar también la necesidad de ampliar la información sobre la posibilidad de incluir secuencias genéticas en bases de datos, su finalidad y las políticas para su gestión, incluso si se va a proceder a la anonimización. En cuanto a la realización de análisis clínicos de secuenciación de genomas completos, como se ha señalado, los puntos recogidos en la LIB parecen suficientes, en el sentido de que proporcionan la información adecuada para que el sujeto sea consciente de las características de estos análisis, siempre que se explique que se va a realizar esta técnica en particular.

Uno de los puntos más conflictivos y discutidos en relación con el régimen de la donación de material humano con fines de investigación es el que se refiere a la amplitud del consentimiento. Si el consentimiento para el tratamiento de datos de carácter personal debe ser específico, como también debe serlo para la participación en una investigación biomédica, ¿es legítimo que un sujeto done su muestra biológica para “cualquier finalidad dentro del marco de la investigación biomédica”? El artículo 70.2 de la LIB ha sentado las bases para el establecimiento de un régimen flexible cuando se articulan garantías reforzadas en la gestión de las muestras: la LIB alude al consentimiento con fines “de investigación” (consentimiento amplio).

En efecto, en España se ha articulado un sistema que ofrece distintas posibilidades y que trata de conciliar los distintos intereses, del donante, de los investigadores y de la sociedad en general como beneficiaria última del avance de la investigación biomédica: es legítimo el consentimiento en términos muy amplios si la gestión de la muestra se lleva a cabo en unas condiciones que garantizan la efectividad de los derechos del sujeto en relación con sus muestras y datos mediante controles muy estrictos por parte de una institución responsable, que actúa como “un intermediario” entre el donante y el investigador.

Así, las muestras se utilizarán en un proyecto concreto, en una colección o en un biobanco, que son tres regímenes diferenciados tal como prevé la letra c del artículo 23 del RD. En el último caso, el consentimiento se puede otorgar para cualquier investigación biomédica, lo que se justifica por el control encomendado a estos establecimientos y por las garantías que suponen las condiciones para su autorización y funcionamiento. La amplitud del consentimiento se irá reduciendo conforme estos requisitos van disminuyendo. El consentimiento se puede referir a varios de estos fines siempre que se cumplan las condiciones propias de todos ellos.

Por otra parte, junto a este régimen general existe uno transitorio y otro excepcional con requisitos similares pero no idénticos:

Según la Disposición transitoria segunda de la LIB, referida a muestras almacenadas con anterioridad, "las muestras biológicas obtenidas con anterioridad a la entrada en vigor de esta ley podrán ser tratadas con fines de investigación biomédica cuando el sujeto fuente haya dado su consentimiento o cuando las muestras hayan sido previamente anonimizadas. No obstante, podrán tratarse muestras codificadas o identificadas con fines de investigación biomédica sin el consentimiento del sujeto fuente, cuando la obtención de dicho consentimiento represente un esfuerzo no razonable en el sentido que se indica en el párrafo i) del artículo 3 de esta ley, o no sea posible porque el sujeto fuente hubiera fallecido o fuera ilocalizable. En estos casos se exigirá el dictamen favorable del comité de ética de la investigación correspondiente, el cual deberá tener en cuenta, como mínimo, los siguientes requisitos:

- a) que se trate de una investigación de interés general;
- b) que la investigación sea menos efectiva o no sea posible sin los datos identificativos del sujeto fuente;
- c) que no conste una objeción expresa del mismo, y
- d) que se garantice la confidencialidad de los datos de carácter personal".

En cuanto al régimen excepcional, la LIB incorporó en su artículo 58 lo previsto en la Recomendación (2006) 4 (añadiendo algún requisito):

"2. El consentimiento del sujeto fuente será siempre necesario cuando se pretendan utilizar con fines de investigación biomédica muestras biológicas que hayan sido obtenidas con una finalidad distinta, se proceda o no a su anonimización.

No obstante lo anterior, de forma excepcional podrán tratarse muestras codificadas o identificadas con fines de investigación biomédica sin el consentimiento del sujeto fuente, cuando la obtención de dicho consentimiento no sea posible o represente un esfuerzo no razonable en el sentido del artículo 3.i de esta Ley. En estos casos se exigirá el dictamen favorable del CEI correspondiente, el cual deberá tener en cuenta, como mínimo, los siguientes requisitos:

- a) que se trate de una investigación de interés general;
- b) que la investigación se lleve a cabo por la misma institución que solicitó el consentimiento para la obtención de las muestras. [Este requerimiento, no recogido en la Recomendación de 2006, es el que más polémica puede suscitar, tanto por la dificultad de concretar el alcance del término "institución", como por la limitación que puede suponer para intercambiar muestras y datos].

- c) que la investigación sea menos efectiva o no sea posible sin los datos identificativos del sujeto fuente;
- d) que no conste una objeción expresa del mismo; y
- e) que se garantice la confidencialidad de los datos de carácter personal”.

Estos requisitos deberán ser evaluados conjuntamente, ser ponderados entre sí y adaptados al tipo de muestras o datos que se pretenda utilizar sin consentimiento. Si se trata únicamente de datos (p. ej., datos de la historia clínica) recordemos que el consentimiento del sujeto no es necesario si se van a anonimizar. Se debe insistir en que el acceso a los datos antes de su anonimización se llevará a cabo por quien esté legitimado para ello. El investigador recibirá los datos una vez se hayan anonimizado.

DERECHO DE ACCESO

Este derecho de acceso es distinto del derecho a la información relevante para la salud del sujeto o de sus familiares obtenida en investigación, supuesto en que dicho sujeto se posiciona como paciente y no ya como donante. Tampoco se corresponde con el derecho de acceso a los resultados generales de la investigación, que no son datos de carácter personal (art. 27.2 de la LIB).

El derecho de acceso a los datos de carácter personal obtenidos en investigación es una proyección del derecho de acceso que la LOPD reconoce al sujeto en relación con cualquier dato objeto de tratamiento. Abarca cualquier dato genético, incluyendo datos crudos, datos filtrados y analizados (sin que se pueda limitar a los datos genéticos validados y relevantes para la salud a los que concretamente se ha referido el artículo 32 del RD 1716/2011), etc., aunque ya se sugirió la posibilidad de establecer ciertas precauciones (v. pág. 76).

No cabe renuncia al derecho y, por consiguiente, no cabe plantear una investigación en la que no se contemple su ejercicio. Como se dijo, si los datos se almacenan por un prestador de servicios (encargado del tratamiento), deben articularse medidas que permitan el ejercicio del derecho por el responsable (p. ej., si se contratan servicios de almacenamiento en la nube).

DERECHO A LA CANCELACIÓN DE DATOS

La LIB recoge con carácter general la facultad de revocación del consentimiento para investigación (art. 4.3) sin necesidad de aducir motivo alguno, y de forma específica en relación con el uso de muestras biológicas que, lógicamente, solo se podrá ejercer si la muestra no ha sido anonimizada. Así, el artículo 60.3 establece que “el consentimiento podrá ser revocado, totalmente o para determinados fines, en cualquier momento. Cuando la revocación se refiera a cualquier uso de la muestra, se procederá a su inmediata destrucción, sin perjuicio de la conservación de los datos resultantes de las investigaciones que se hubiesen realizado con carácter previo”.

Estos efectos deben estar descritos en la información previa al consentimiento inicial para asegurar que el sujeto conoce las consecuencias que tendrán sus decisiones en el futuro: “El documento de consentimiento para la obtención, almacenamiento o conservación y utiliza-

ción de muestras biológicas de origen humano con fines de investigación biomédica contendrá como mínimo la siguiente información para el sujeto fuente [...] k) Derecho de revocación del consentimiento, total o parcial, a ejercer en cualquier momento, y sus efectos, incluida la posibilidad de la destrucción o de la anonimización de la muestra y de que tales efectos no se extenderán a los datos resultantes de las investigaciones que ya se hayan llevado a cabo” (art. 23.2 del RD).

La revocación del consentimiento supone que la muestra ya no se someterá a nuevos análisis o que será anonimizada, pero sus efectos no se extenderán a los datos resultantes de las investigaciones que ya se hayan llevado a cabo (art. 59.e de la LIB).

El interés de los investigadores en relación con los resultados de su actividad profesional debe ser reconocido y protegido, y por consiguiente la revocación del consentimiento no supondrá la supresión o eliminación los datos. Además, los datos deben conservarse para atender a eventuales responsabilidades derivadas del tratamiento. Estas exigencias no pueden suponer una vulneración del derecho de los sujetos a su intimidad y control de su información. Por eso, si el consentimiento se revoca, los datos quedarán “bloqueados” (solo se utilizarán para atender aquellas responsabilidades) y solo podrán publicarse o cederse de manera agregada sin que sea posible la identificación de los sujetos. En este sentido, aunque el consentimiento es necesario para legitimar el tratamiento de los datos en investigación, los efectos de su revocación se corresponden más bien con los del ejercicio del derecho de cancelación tal como se concibe en la LOPD (v. pág. 77 donde se explicaban las diferencias entre el derecho de revocación del consentimiento y el derecho de cancelación).

Obviamente, si los datos se han sometido a un proceso de anonimización, no sería posible ejercer este derecho.

DERECHO A LA INFORMACIÓN SOBRE LOS RESULTADOS Y DERECHO A NO SABER

A pesar de que en el contexto de la investigación la finalidad de los análisis no es el beneficio directo para el sujeto participante, también se reconoce el derecho a la información sobre los resultados obtenidos cuando estos pueden tener trascendencia para la salud del sujeto fuente o sus familiares. Se exige además la disponibilidad de consejo genético si estuviese indicado. Se reconoce también el derecho a no saber y se plantean cuestiones parecidas a las explicadas en un apartado anterior (v. pág. 78) donde se analizó la regulación de este derecho en el ámbito clínico.

EL SUJETO MENOR

Como ocurre en general con la investigación biomédica, la investigación con muestras y datos de menores está sujeta a unas condiciones particulares que tratan de garantizar la protección de una población especialmente vulnerable. Estas condiciones se enumeran en el artículo 58 de la LIB, según el cual la investigación debe quedar justificada, es decir la utilización de muestras de menores debe ser necesaria para la investigación concreta (58.5.b y c); el riesgo de la intervención invasiva, caso de ser precisa, no ha de superar el mínimo (art. 58.a); y se tendrá en cuenta la voluntad del menor, o incluso si su madurez es suficiente será él mismo quien consienta (art. 58.d de la LIB).

El investigador o el responsable del biobanco tienen la responsabilidad de que el menor se involucre en la toma de decisión en función de su madurez y que sea informado de forma adecuada de lo que supone su participación en el presente y en el futuro. Además, los representantes legales deben ser informados de la necesidad de su participación en la efectividad de la garantía prevista en el artículo 23.2.n del RD 1716/2011 (cuando el sujeto alcance la mayoría de edad será él quien ejerza los derechos correspondientes sobre las muestras almacenadas y los datos asociados).

Un asunto discutido en relación con la obtención de datos y muestras de menores con fines de investigación biomédica es si es o no necesario contactar con el sujeto para solicitar un nuevo consentimiento cuando alcance la mayoría de edad, especialmente cuando su utilización se pueda prolongar durante largos periodos, incluso indefinidamente en régimen de colección o biobanco.

A favor de esta exigencia se sostiene que representa el único mecanismo para garantizar el derecho del propio sujeto a decidir sobre su participación en la investigación con muestras o datos, que en el caso de datos genómicos tiene una especial relevancia por tratarse de información que puede revelar información inesperada y que podrá suponer riesgos más elevados de identificación del sujeto. El consentimiento dado por los representantes se otorga en un momento concreto y legitima la toma de decisiones que afecten al menor en ese aspecto. En el almacenamiento de datos genéticos, dicha decisión, tomada en su momento, podría afectar a ese menor ya adulto en el futuro, sin él saberlo. Quien tendría la responsabilidad de dar tal información sería el titular del registro o fichero donde se encuentra almacenada, a fin de dar al sujeto fuente la posibilidad de ejercer sus derechos. Tal información requeriría contactar con el interesado, sin necesidad de esperar a que esos datos deban ser utilizados.

En contra, puede decirse que desde el punto de vista jurídico el consentimiento que prestaron los representantes legales en lugar del menor sigue siendo válido alcanzada la mayoría de edad, que la obtención de muestras o datos no significa su efectiva utilización en una investigación y que, en todo caso, cuando se vayan a utilizar, un comité de ética deberá decidir en el supuesto concreto si es necesario el consentimiento específico del sujeto (artículo 32.2 del RD 1716/2011). Además, la utilización de los datos y muestras deberá estar siempre sujeta a las limitaciones propias de la investigación pediátrica.

En todo caso, el investigador o el responsable del biobanco deberán tener en cuenta la fecha en que el sujeto alcance la mayoría de edad a efectos de futuros contactos (p. ej., para informar sobre el cierre del biobanco, para comunicar un hallazgo relevante para la salud, etc.), y también de que esté a su disposición la información prevista en el artículo 32 (art. 32.3: “En el caso de utilización de muestras de menores de edad con fines de investigación biomédica, según lo previsto en el artículo 58.5 de la Ley 14/2007, de 3 de julio, el biobanco y las personas responsables de la colección o del proyecto de investigación tendrán la información a la que se refiere el apartado 1 de este artículo a disposición de la persona representante legal del sujeto fuente hasta que este alcance la mayoría de edad, y del propio sujeto fuente a partir de ese momento”).

El Comité de Bioética del Consejo de Europa, en sus trabajos sobre la Revisión de la Recomendación 4 (2006) relativa a la utilización de muestras en investigación biomédica, prevé que

cuando el menor donante alcance la mayoría de edad debería ser recontactado (arts. 12 y 14.4). La repercusión de esta eventual posición, en su caso, debería consensuarse para que no se genere inseguridad para los biobancos y los investigadores.

Finalmente, recuérdese que en algunos casos la investigación debe ser puesta en conocimiento del Ministerio Fiscal. La Circular 8 /2011 de la Fiscalía General del Estado de 16 de noviembre sobre criterios para la unidad de actuación especializada del Ministerio Fiscal en materia de protección de menores señala que esta obligación de comunicación se limita a los casos en que se vaya a someter al menor a un procedimiento invasivo y no se vayan a producir resultados de beneficio directo para él. No se exige la puesta en conocimiento del fiscal de cada otorgamiento del consentimiento con carácter previo, sino una descripción en conjunto de la investigación. Por ello, no es necesario comunicar a la Fiscalía ni la identidad de los menores ni enviar el documento de consentimiento.

INVESTIGACIÓN CON MUESTRAS O DATOS DE FALLECIDOS

El RD 1716/2011 ha desarrollado el régimen de utilización de muestras de fallecidos con fines de investigación en el artículo 26: se atiende en primer lugar a la voluntad expresada en vida, para cuya comprobación se prevé la indagación de la existencia de instrucciones previas, y en ausencia de estas, la consulta a los familiares más próximos y a los profesionales que le atendieron en el centro sanitario (se entiende, en el centro donde el sujeto falleció).

Si la voluntad expresa a favor o en contra de la donación no se acreditara, se establece una presunción de donación sustentada en los principios que en nuestro ordenamiento fundamentan la utilización de partes del cuerpo de personas fallecidas en beneficio de terceros y que “refieren al respeto a la voluntad del sujeto en vida y a la solidaridad y participación ciudadana en el progreso científico” (exposición de motivos del RD).

Por otra parte, el artículo 26 dispone que “Las personas vinculadas a la persona fallecida por razones familiares o análogas podrán dirigirse a los responsables de los ficheros o de los tratamientos que contengan datos de dicha persona fallecida con la finalidad de notificar el óbito, aportando acreditación suficiente del mismo, y solicitar, justificando la concurrencia de un interés relevante, la cancelación de los datos o la anonimización de las muestras”. Esta previsión está estrechamente relacionada con el artículo 2 del Real Decreto 1720/2007, de 21 de diciembre, por el que se aprueba el Reglamento de desarrollo de la Ley Orgánica 15/1999, de 13 de diciembre, de protección de datos de carácter personal, que recoge esta misma posibilidad en general tras establecer que el reglamento no será de aplicación a los datos referidos a personas fallecidas.

En cuanto a los datos genéticos que ya hubieran sido obtenidos en vida, no se exigen estas indagaciones si se van a anonimizar. En caso contrario, sería razonable aplicar el régimen descrito para las muestras que, además, se inspiró en parte en el Reglamento de Desarrollo de la LOPD.

Los familiares del sujeto de la investigación

Véase página 74 y lo señalado en relación con sus facultades en el apartado anterior cuando se trata de familiares de fallecidos.

El investigador principal, el equipo de investigación y su institución

El análisis de los datos de secuenciación genómica con fines de investigación se lleva a cabo en centros de investigación (universidades, hospitales, centros públicos y empresas privadas de secuenciación,...) que incluyen esta actividad entre sus objetivos. El responsable de los datos a efectos de la LOPD será la institución que albergue el o los ficheros (v. pág. 62).

Tanto el investigador responsable del proyecto como el resto del equipo son usuarios autorizados que tratan los datos para desempeñar las funciones propias de su función en dichas instituciones.

Con frecuencia ocurre que las muestras o los datos provienen de un centro distinto al del investigador: por ejemplo se transfieren de un hospital a un centro de investigación básica o de un biobanco a diferentes investigadores. En estos casos, las muestras y los datos se pueden transferir de forma anonimizada si es que no se precisa la trazabilidad con la identidad del sujeto. Sin embargo, en casi todos los estudios es necesario que exista este vínculo entre muestra secuenciada e información clínica, ya que la interpretación de los datos derivados de la secuenciación, aunque sea en investigación básica, suele requerir el acceso a los datos clínicos para identificar genes o mutaciones relacionadas con algunos parámetros clínicos (como supervivencia, respuesta al tratamiento,...). Los datos y las muestras se transfieren entonces de forma reversiblemente dissociada. El centro de origen de los datos o muestras es el responsable del fichero que alberga y le corresponde obtener el consentimiento del titular de los datos o sujeto fuente para su cesión con fines de investigación; el centro receptor de los datos o muestras codificadas también adquiere responsabilidades en relación con el almacenamiento y utilización de los datos y el material en los términos descritos en el capítulo 4. Será preciso concertar los correspondientes acuerdos y establecer medidas proporcionales para garantizar que los derechos de los sujetos se respetan por parte de los cesionarios: los ficheros de datos crudos y procesados se deben almacenar en servidores seguros con acceso controlado. Dichos ficheros no deben contener información personal ni clínica del individuo estudiado, tiene que existir un sistema de recodificación de las muestras, y el acceso a dicho fichero solo deberían tenerlo un número muy pequeño de personas.

Por lo que se refiere al “responsable del fichero de datos del biobanco” (art. 66 de la LIB), es una figura que forma parte de la estructura de esta institución, junto con el titular y el director científico. Su función es atender “las solicitudes de ejercicio de los derechos de acceso, rectificación, cancelación u oposición formuladas por los participantes, de conformidad con lo dispuesto en la normativa vigente sobre protección de datos de carácter personal”. Sin embargo, se debe recordar que es el titular quien ostenta la responsabilidad del biobanco y por lo tanto decide sobre la finalidad del fichero de datos. Por consiguiente, figurará como responsable en el registro y será quien asuma las responsabilidades correspondientes, incluida la de garantizar el ejercicio de los derechos de los sujetos. Parece desprenderse de la redacción de la LIB que, además del titular (responsable a efectos de la LOPD), otra persona gestionará las peticiones de acceso y cancelación de los datos, pero en función de las definiciones y obligaciones descritas en la LOPD, se tratará de un usuario autorizado, que hará efectivas las responsabilidades del titular del biobanco, pero no las asumirá.

Otras instituciones participantes. Posibilidad de proyectos internacionales

La colaboración científica internacional representa un pilar básico de la investigación. La investigación genómica es un paradigma de esta realidad. Debido a la dificultad de conseguir un gran número de muestras de determinadas patologías, y a que la mayor parte de los análisis de investigación se basan en la identificación de cambios de manera estadística (por lo que a mayor número de muestras mayor fiabilidad de los resultados), las investigaciones genómicas constituyen una de las áreas de colaboración más dinámica en la actualidad. A esto hay que añadir que el envío de muestras de ADN a otros países para su secuenciación con fines de investigación y/o diagnóstico clínico es cada vez más frecuente: el coste de secuenciación es muy distinto según los países donde se desarrolle, incluso utilizando la misma tecnología. De hecho, la secuenciación en el extranjero puede suponer ahorros de entre el 50% y el 70% del coste comparado con los centros nacionales.

El desarrollo previsto en la LIB en relación con la circulación internacional de muestras, ya sea para proyectos conjuntos con centros extranjeros, ya sea para atender solicitudes de investigadores extranjeros, todavía no se ha llevado a cabo. Por consiguiente, en esta materia deben tenerse en cuenta las siguientes cuestiones:

- A. En la regulación de la LOPD, las comunicaciones de datos a los estados miembros de la Unión Europea o del Acuerdo sobre el Espacio Económico Europeo no se consideran transferencias internacionales de datos, sino una modalidad más de cesiones de datos. Ello es así porque la Directiva 95/46/CE establece un sistema armonizado de garantías, uno de cuyos objetivos principales es garantizar el libre flujo de datos para un adecuado desarrollo del mercado único. Por tanto, solo serán exigibles los requisitos establecidos para una cesión de datos como si se realizara dentro del territorio español.

En el caso de que la cesión de datos se realice a terceros países distintos de los señalados caben varias opciones.

La primera de ellas es la obtención de un consentimiento específico e informado del sujeto fuente para llevarla a cabo. Esta alternativa puede considerarse la más sencilla en la práctica si se tiene en cuenta la pluralidad de aspectos sobre los que aquel ha de prestar el consentimiento. Así, la colaboración entre centros para desarrollar un proyecto o la cesión de muestras o datos para proyectos en el extranjero debe formar parte de la información que reciba el sujeto participante antes de otorgar su consentimiento, donde se identificarán los centros que van a participar, o se advertirá de la disponibilidad de esta información. Jurídicamente, esto significa que el sujeto consiente para la cesión desde la institución donde se recogen sus datos hacia las otras instituciones participantes en el proyecto y en el marco de esa finalidad específica.

No obstante, la LOPD permite realizar transferencias internacionales de datos sin el consentimiento de los interesados en relación con las mismas, siempre que concurran otras garantías.

Los principales supuestos son los siguientes:

- Transferencias internacionales de datos a terceros países sobre los que la Comisión Europea ha adoptado una Decisión de adecuación, que implica el reconocimiento de

que ofrecen un sistema de garantías equivalentes al de la Unión Europea y no exigen autorización de la Agencia Española de Protección de Datos (AEPD).

- Hasta la Sentencia del Tribunal de Justicia de la Unión Europea de 6 de octubre de 2015 estaban legitimadas las transferencias internacionales a entidades norteamericanas adheridas al denominado “Acuerdo de Puerto Seguro”. Sin embargo, la citada sentencia ha invalidado con efectos retroactivos la Decisión de Adecuación de la Comisión Europea 2000/520/CE (llamada “Decisión de Puerto Seguro”), por lo que dichas transferencias no podrán ampararse en ella y deberán fundarse en otras modalidades de garantías.
- Transferencias internacionales a terceros países que no ofrecen un nivel de protección adecuado, que deben ser autorizadas por el director de la AEPD.

El supuesto más habitual de este tipo de transferencias son las que ofrecen las garantías recogidas en las “cláusulas contractuales tipo” aprobadas por la Comisión Europea, bien para las transferencias de responsable del tratamiento a otro responsable en el tercer país, bien las específicas para la transferencia de un responsable del tratamiento en España a un encargado del tratamiento en el país tercero. (Puede obtener más información sobre los países que tienen reconocido un nivel adecuado de protección o sobre los requisitos de las transferencias internacionales de datos en: https://www.agpd.es/portalwebAGPD/canalresponsable/transferencias_internacionales/index-ides-idphp.php³⁷).

- B. Se puede plantear si se pueden llevar a cabo transferencias internacionales de datos genéticos considerados de carácter personal sin consentimiento *inequívoco* del titular cuando lo autorice el director de la Agencia de Protección de Datos (art. 33 de la LOPD) y si esta autorización es necesaria cuando la transferencia ha sido informada positivamente por el comité de ética correspondiente. En primer lugar, teniendo en cuenta que la regulación específica (LIB y RD 1716/2011) prevé la necesidad de intervención de un comité de ética en estos casos, parece que sin esta evaluación (positiva) la transferencia no podría llevarse a cabo. En cuanto a la segunda cuestión, el dictamen favorable del comité, si bien certifica que lo previsto en la LIB se cumple efectivamente en un determinado contexto, no será suficiente a efectos de la LOPD y por consiguiente se deberá solicitar la autorización del director de la Agencia.
- C. En las cesiones internacionales se deberán articular las garantías necesarias para asegurar que los derechos reconocidos a los donantes por el ordenamiento jurídico español podrán también ser efectivos si la investigación se lleva a cabo en otro país. El acuerdo de cesión de muestras y/o datos entre los centros debe contemplar este compromiso, cosa que también ocurre en las cesiones no consideradas internacionales.

La adopción de medidas de seguridad en el tratamiento de los datos cuando el país de destino no forme parte del espacio económico europeo o no es considerado como *safe harbour* tiene una trascendencia especial en este caso. En el acuerdo debería figurar:

- Finalidad para la que se cede la muestra.
- Obligaciones del cesionario:
 - Utilización de la muestra con la finalidad exclusiva para la que se envía.

- o No identificación de los donantes.
 - o Aplicación de medidas para asegurar la trazabilidad de la muestra.
 - o Aplicación de medidas de seguridad a los datos y concreción de las mismas.
 - o Garantía de la disponibilidad de la información genética validada y relevante para la salud que, en su caso, se obtenga del análisis de las muestras.
 - o Compromiso de destrucción o devolución del material sobrante una vez finalizado el proyecto.
 - o Compromiso de hacer efectiva la revocación del consentimiento en caso de que se produzca.
 - Obligaciones del cedente:
 - o Tipo y cantidad de muestra que se compromete a enviar.
 - o Condiciones del transporte.
 - Otras cuestiones:
 - o Condiciones relativas a la propiedad intelectual o industrial de los resultados de la investigación. Consecuencias del incumplimiento de los términos del acuerdo.
 - o Jurisdicción aplicable o sometimiento a otro procedimiento de solución de controversias. (Téngase en cuenta que la jurisdicción no es disponible contractualmente en lo que afecte al tratamiento de datos personales).
- D. Por otra parte, el responsable del fichero deberá notificar la previsión de transferencias internacionales de datos al Registro General de Protección de Datos (art. 20.2.e de la LOPD) cuando el fichero se inscriba o posteriormente si es que entonces no hubiera estado previsto.

Finalmente, debe recordarse que cuando los datos no se consideran de carácter personal por no preverse la trazabilidad con la identidad del sujeto, este régimen no es aplicable, si bien en el caso de secuencias genéticas no debe descartarse la aplicación de medidas para datos no considerados de carácter personal, encaminadas a garantizar precisamente que no se produzca la identificación.

Otro escenario es el que se plantea si el investigador requiere la prestación de un servicio por un tercero no miembro del equipo de investigación, que será considerado “encargado del tratamiento” según la LOPD (v. pág. 59). La transferencia de datos o muestras debe igualmente formalizarse en un acuerdo entre el responsable del fichero y el tercero prestador de servicios. Si bien la LOPD no exige consentimiento del titular de los datos en estos supuestos, la información de que el análisis se ha “externalizado” deberá estar disponible para el donante tal como prevé la LIB (art. 59). Tras el servicio propiamente de secuenciación, el investigador obtiene, bien las secuencias crudas para su análisis posterior (generalmente los datos de secuenciación se almacenan en servidores de proveedor de servicios y pueden ser descargadas de manera segura por el investigador) o bien un análisis completo que contrata y se realiza en el mismo centro que oferta el servicio de secuenciación.

LA GESTIÓN DE LAS BASES DE DATOS GENÉTICOS A DISPOSICIÓN DE LA COMUNIDAD CIENTÍFICA

Las bases de datos de resultados de secuenciación más abundantes son aquellas que recogen todas las variantes halladas y en las que la identidad del sujeto secuenciado está disociada (de manera irreversible o reversible). Se utilizan principalmente con fines estadísticos o epi-

demográficos, ya que recogen la variabilidad genética (variantes neutras o deletéreas) de una población concreta. Este tipo de bases son empleadas frecuentemente para realizar filtrados según MAF (*minimum allele frequency*). Las más consultadas y completamente públicas son la base Exome Variant Server (que alberga más de 6500 exomas), la base de datos del Exome Aggregation Consortium (ExAC), que alberga datos de más de 60 000 exomas, o la base de datos del proyecto 1000 Genomes.

En el ámbito nacional cabe destacar la base que está siendo creada por la Plataforma bioinformática de enfermedades raras (CIBERER Exome Server), que pretende ofrecer información sobre la variabilidad en la población española y está actualmente accesible en una primera versión beta.

También existen bases de datos que recogen mutaciones relacionadas con patología o que predisponen al desarrollo de enfermedades de base genética como *The Human Gene Mutation Database* (que es solo parcialmente pública) o a nivel nacional la recientemente presentada SpainMDB (*Spain Mutation DataBase*). Este tipo de bases de datos, que podrían generarse en cada centro hospitalario, podrían ser compartidas o incluso públicas para uso de la comunidad científica y médica. Su implementación sería relativamente sencilla porque ya existen plataformas libres adaptables (*open source*) como la *Leiden Open Variation Database* creada en la Universidad de Leiden (Holanda). Por lo tanto, los costes de desarrollo inicial son muy bajos y los gastos fundamentalmente serían del personal que realice las tareas de administración y mantenimiento de la base y de la infraestructura informática que la soporte.

Por otra parte, existen repositorios donde los investigadores pueden enviar los datos de secuenciación y otra información asociada para hacerla disponible a toda la comunidad científica a través de distintos procedimientos (EGA, dbGAP,...).

El acceso a todas estas bases de datos varía según su potencial identificador. Cuando se considera que no existe en absoluto, los datos son de acceso libre (*open access*). En otro caso, se toman precauciones (y se deben adoptar garantías jurídicas para la no reidentificación, como la previsión de medidas en caso de incumplimiento) y se controla el acceso (*controlled access*).

En efecto, en este segundo supuesto, el depósito de los datos se debe hacer de una manera segura, y todos los ficheros depositados deben estar encriptados para evitar que, en caso de incidentes, se pueda obtener información de ellos. El interesado en el acceso debe enviar una petición para que sea evaluada por un DACO (*Data Access Compliance Office*), que si la acepta permitirá la descarga directa de los datos crudos. La autorización para el acceso a los datos se concederá para una única persona que no solo tendrá una contraseña para el acceso, sino también una llave de seguridad física (tipo RSA SecurID). Estas medidas de seguridad garantizarán que la usurpación de la contraseña, o la de la llave física, no sea suficiente para acceder a los datos y dar acceso a personas no autorizadas, sino que será necesario tener ambos elementos.

A nivel internacional han surgido estrategias como Global Alliance, que suponen una gran oportunidad para desarrollar protocolos de intercambio de datos genómicos así como de datos clínicos de pacientes, por lo que será interesante estudiar los problemas encontrados y soluciones adoptadas en este proyecto piloto.

Los sujetos de los datos deben conocer estas políticas. Algunos de los proyectos puestos en marcha en los últimos años, como el International Cancer Genome Consortium (ICGC), tenían ya durante su fundación este espíritu colaborativo, lo que condujo a los distintos países participantes a redactar los documentos de consentimiento informado para los pacientes en unos términos que permitiesen este grado de colaboración imprescindible para el avance científico, con el establecimiento de comités de acceso a los datos y medidas de control. En el caso de proyectos más modestos, la colaboración es mucho más difícil.

Los comités evaluadores no son propiamente comités de ética, ni comités de biobancos, sino que se trata de un grupo de expertos implicados en los proyectos con conocimientos suficientes para llevar a cabo esta función. No obstante, los comités de ética acreditados deben intervenir evaluando estos procedimientos.

La institución que alberga la base de datos no tiene la categoría de encargado o responsable del tratamiento porque los datos no se consideran de carácter personal ya que no existe trazabilidad con el sujeto fuente. Sin embargo, se deberían establecer acuerdos entre las partes que garanticen el uso legítimo de los datos, la continuidad de la supervisión y que los sujetos no serán identificados en el futuro (v. cap. 4).

Por otro lado, posiblemente sea interesante crear bases de datos no disociadas en el ámbito local (principalmente para los casos que no han podido ser cerrados con la información disponible en un momento determinado). Esto exigiría un desarrollo informático y una planificación previa que tenga en cuenta este hecho y que permitiera aplicar las medidas necesarias para garantizar la seguridad de los datos: una sólida política de control y registro de los accesos, adopción de garantías jurídicamente vinculantes para la no reidentificación, encriptación de datos y diferentes niveles de seguridad tanto físicos como de *software* (en un nivel incluso más elevado que el previsto para las bases de datos referidas anteriormente). En el plano más práctico habría que identificar, según la estructura de funcionamiento de la base de datos, a quién correspondería depositar datos y quién se encargaría de su administración, custodia y de la hipotética cesión de datos a otras instituciones (responsable del fichero). Se debería tener en cuenta que la posibilidad de cesión de muestras de sujetos identificables obtenidas en proyectos concretos es muy limitada, puesto que requiere consentimiento para cada una de las transferencias. Si se aplica este criterio a los datos obtenidos será imposible su incorporación a este tipo de archivos. Además, en general, aunque puede haber excepciones, los datos ni se originan ni se almacenan en un biobanco, que no analiza las muestras, sino que las cede a proyectos. En definitiva, las bases de datos deben nutrirse de resultados de proyectos concretos, pero la aplicación del criterio de limitación en la cesión de muestras, que solo se plantea de modo amplio en régimen de donación para biobanco, dificulta esta posibilidad. Se debería reflexionar sobre el sentido de esta limitación y si su justificación cuando se trata de muestras es trasladable en los mismos términos a los datos resultantes de los proyectos ante el escenario científico y tecnológico que se presenta. En cualquier caso, se debería promover que los biobancos, que son establecimientos con características idóneas para garantizar los derechos de los sujetos y promover la disponibilidad de recursos para la investigación, desempeñaran un papel más relevante en este ámbito.

En otro orden de cosas, se requerirá necesariamente la contratación de personal cualificado para realizar estas tareas. Los recursos materiales deberán incluir sistemas de almacenamiento

en red (NAS, *network-attached storage*) convenientemente configurados, que usen protocolos de comunicación seguros y que sean escalables. También será necesario garantizar la integridad física de los equipos en un recinto acondicionado para tal uso (climatización autónoma, sistemas antiincendios y control de acceso).

En la actualidad, este problema no es muy importante, ya que la cantidad de genomas secuenciados disponibles es limitada y son pocos laboratorios los que tienen capacidad de análisis suficiente, pero a medida que se vaya generalizando el empleo de estas tecnologías, la necesidad de compartir información será cada vez mayor, e incluso en el caso de secuencias con finalidad clínica, la posibilidad de pedir una segunda opinión basándose en los mismos datos genéticos, tanto a entidades públicas como privadas, podrá convertirse en un proceso habitual. Así, aunque no constituya una necesidad acuciante, sería interesante proponer mecanismos que permitan el intercambio de datos genómicos y faciliten el avance científico, preservando los derechos de los donantes.

Una posibilidad podría ser el simplificar el proceso de firma de convenios entre los centros implicados: elaborar modelos adaptables a situaciones concretas y que sean supervisados por comités de ética, que revisen también los procedimientos de gestión. Otra medida interesante sería estudiar la posibilidad de fomentar el reconocimiento de centros que adopten determinadas políticas de seguridad, información y transparencia para facilitar su participación en estos flujos de datos. Se debería determinar a qué entidad podría corresponder esta acreditación y según qué estándares. Esta cuestión debe ser objeto de un análisis más específico y minucioso. La adhesión a foros como la Global Alliance y sus códigos de conducta, a lo que ya se ha hecho referencia, puede ser un primer paso en este sentido. Además, este tipo de mecanismos es posible que exijan completar y unificar la información que se da a los pacientes o sujetos de la investigación.

De manera complementaria al establecimiento de procedimientos que permitan acceder y compartir datos genómicos, uno de los objetivos que debería ser prioritario es iniciar una labor de formación en bioética para los distintos agentes implicados en este tipo de investigaciones. Así, aunque el personal de los centros sanitarios y los biobancos suele conocer bastante bien la legislación referente al manejo de datos genómicos, la generalización de la secuenciación genómica a laboratorios de investigación supone que gran parte de los profesionales implicados en la generación, manejo y análisis de este tipo de datos nunca haya recibido formación referente a la legislación y consideraciones bioéticas. Esta situación, junto al previsible aumento de profesionales con acceso a este tipo de datos en los próximos años, no hace sino incrementar el riesgo de errores. Por esto es muy recomendable la oferta de cursos o seminarios, comenzando por estudios especializados como pueden ser másters en Bioinformática, Biología Computacional, Genética, etc.

Finalmente, se debería valorar la implicación de la Agencia de Protección de Datos en la gestión de estos ficheros, y en la posibilidad de asesoramiento sobre los procedimientos de gestión. Además, puede ser relevante también su intervención a efectos de autorización de tratamiento de datos que permitan la trazabilidad con la identidad del sujeto pero sin consentimiento, según lo dispuesto en el artículo 50.3 de la LILB: “en casos excepcionales y de interés sanitario general, la autoridad competente, previo informe favorable de la autoridad en materia de protección de datos, podrá autorizar la utilización de datos genéticos codificados, siempre ase-

gurando que no puedan relacionarse o asociarse con el sujeto fuente por parte de terceros". No se debe descartar que una autorización pueda legitimar la cesión de datos genéticos, por ejemplo, para crear registros de enfermedades si se entiende que la investigación científica puede considerarse un "interés sanitario general".

RESUMEN DEL CAPÍTULO

En la obtención, almacenamiento y cesión de los datos genéticos resultantes de NGS se deben aplicar las reglas previstas para la gestión de estos datos en general, si bien algunos aspectos cobran una especial relevancia.

Las indicaciones para la práctica clínica de los análisis han sido reguladas recientemente.

En cuanto al consentimiento de los sujetos, se debe incluir la previsión de secuenciación completa del genoma y valorar la necesidad de ampliar la información previa en relación con la posibilidad de incluir secuencias genéticas en bases de datos, su finalidad y las políticas de gestión, incluso si se va a proceder a su anonimización. Por otra parte, se debe tener en cuenta existe que la posibilidad de hallazgos inesperados. Por esta razón es muy recomendable restringir el uso de NGS en menores a finalidades diagnósticas o de investigación no predictivas y focalizar en lo posible el análisis de variantes o genes relacionados con la patología usando los filtros adecuados. También por este motivo el reconocimiento del derecho a no saber cobra una especial trascendencia. Es importante que lo relativo a su ejercicio quede documentado y que se identifique a un familiar o allegado del sujeto para que pueda ser contactado en caso de hallazgos que pudiesen afectar a terceros y aquel haya ejercido su derecho a la ignorancia. El papel de los comités de ética asistencial puede ser muy relevante en este sentido: deberían intervenir para ponderar los intereses en el caso concreto, incluida la posibilidad de que los familiares prefirieran mantenerse en la ignorancia, y para solucionar aspectos concretos que se plantean en esta situación, como por ejemplo quién y cómo se contacta con los familiares.

El almacenamiento de estos datos, considerados como información clínica, debe contemplarse desde la perspectiva del archivo de esta documentación y por lo tanto se deben mantener vinculados a la historia y disponibles para atender la salud del sujeto o de sus familiares. El archivo de esta documentación es responsabilidad del centro sanitario. En el supuesto de que contrate la gestión de la documentación clínica con una entidad externa se deberá formalizar el contrato correspondiente por el que dicha entidad pasará a ser un “encargado del tratamiento”. También deberá cumplir con las obligaciones del encargado del tratamiento la institución que presta un servicio de análisis y que recibe las muestras con esta finalidad.

Cuando se trata de análisis en el contexto de una investigación, los responsables del tratamiento serán igualmente las instituciones o profesionales que incluyen esta actividad entre sus objetivos. Tanto el investigador responsable del proyecto como el resto del equipo son usuarios autorizados que tratan los datos para desempeñar las funciones propias de su función en dichas instituciones. Si el investigador requiere la prestación de un servicio por un tercero no miembro del equipo de investigación, este será considerado encargado del tratamiento.

Por lo que se refiere al derecho de acceso a los datos por parte del titular, en cualquier contexto, se podría contemplar la necesidad de intervención de un profesional para asegurar su comprensión (consejo genético) en lugar de una comunicación directa, o al menos de advertir de su trascendencia.

En cuanto a las posibilidades de disponibilidad de datos genéticos para toda la comunidad científica, variarán según su potencial identificador. Cuando se considera que no existe en

absoluto, el acceso podrá ser libre (*open access*); en otro caso, se deberán tomar precauciones y controlar el acceso (*controlled access*). En este segundo supuesto, todos los ficheros depositados estarán encriptados para evitar que, en caso de incidentes, se pueda obtener información de los mismos. El solicitante de acceso deberá enviar una petición para que sea evaluada por un DACO (*Data Access Compliance Office*), que si la acepta permitirá la descarga directa de los datos crudos. La autorización para el acceso a los datos se concederá para una única persona que no solo tendrá una contraseña para el acceso, sino una llave de seguridad física (tipo RSA SecurID). Estos comités evaluadores no son propiamente comités de ética, ni comités de biobancos, sino que se trata de un grupo de expertos implicados en los proyectos con conocimientos suficientes para llevar a cabo esta función. Los comités de ética acreditados deben intervenir evaluando estos procedimientos.

Por otro lado, posiblemente sería interesante crear bases de datos no disociadas en el ámbito local, lo que exigiría un desarrollo informático y una planificación previa que tuviera en cuenta este hecho y permitiera aplicar las medidas necesarias para garantizar la seguridad de los datos. Se debe tener en cuenta que la posibilidad de cesión de muestras de sujetos identificables obtenidas en proyectos concretos es muy limitada, puesto que requiere consentimiento para cada una de las transferencias. Si se aplicara este criterio a los datos obtenidos sería imposible su incorporación a este tipo de archivos. Por otra parte, en general aunque puede haber excepciones, los datos ni se originan ni se almacenan en un biobanco, que no analiza las muestras, sino que las cede a proyectos. En definitiva, las bases de datos se deben nutrir de resultados de proyectos concretos, pero la aplicación del criterio de limitación en la cesión de muestras, que solo se plantea de modo amplio en régimen de donación para biobanco, dificulta esta posibilidad. Se debería reflexionar sobre el sentido de esta limitación y sobre si su justificación cuando se trata de muestras es trasladable en los mismos términos a los datos resultantes de los proyectos ante el escenario científico y tecnológico que se presenta. En cualquier caso, se debería promover que los biobancos, que son establecimientos con características idóneas para garantizar los derechos de los sujetos y promover la disponibilidad de recursos para la investigación, desempeñen un papel más relevante en este ámbito.

A nivel internacional han surgido estrategias como Global Alliance, que suponen una gran oportunidad para desarrollar protocolos de intercambio de datos genómicos así como de datos clínicos de pacientes, por lo que será interesante estudiar los problemas encontrados y soluciones adoptadas en este proyecto piloto.

Finalmente, debe recordarse que el desarrollo previsto en la LIB en relación con la circulación internacional de muestras, ya sea para proyectos conjuntos con centros extranjeros, ya sea para atender solicitudes de investigadores extranjeros, todavía no se ha llevado a cabo. Por consiguiente, en esta materia la aplicación de los requisitos previstos en la LOPD son especialmente relevantes. Se debería valorar la implicación de la Agencia de Protección de Datos en la gestión de estos ficheros, y en la posibilidad de asesoramiento sobre los procedimientos de gestión.



EFICACIA Y RACIONALIDAD EN LA GESTIÓN DE LAS BASES DE DATOS GENÓMICOS: ORGANIZACIÓN, PLANIFICACIÓN Y SOSTENIBILIDAD

Ya se ha comentado a lo largo de todo este documento que las tecnologías de alto rendimiento están cambiando rápidamente el campo de la detección de enfermedades genéticas. El cambio es tan abrupto, que las nuevas tecnologías también tienen el potencial de detectar enfermedades genéticas incluso antes de que aparezcan los síntomas. Por ello su aplicación responsable tiene un gran impacto en el Sistema Nacional de Salud, así como en la formación y en la preparación de los trabajadores de la salud para decidir y discutir cuándo estas tecnologías son o no son adecuadas para determinadas indicaciones médicas en el plano de la medicina asistencial y en investigación biomédica.

Debemos considerar la relevancia de los resultados en términos de sensibilidad y valor predictivo. Un criterio fundamental para decidir si las pruebas genéticas se deben ofrecer o no en el cuidado regular de la salud es la utilidad clínica, es decir, si la prueba dará lugar a una mejora de las prestaciones de salud de una persona o población y mejorar los flujos de derivación a los servicios correspondientes, que puede y debe ser necesario actualizar.

Todo ello requiere, además, de infraestructuras racionalizadas y personal adecuado a nivel asistencial y de investigación; sobre todo a nivel de la gestión, análisis, almacenamiento, protección y nivel de explotación de los datos genómicos en formato digital.

En este capítulo abordaremos las sugerencias actuales de organización, planificación y sostenibilidad de estas nuevas necesidades, así como también las necesidades de personal y formación y su impacto económico, tanto a nivel de la genética hospitalaria como en infraestructuras de investigación.

MODELOS ORGANIZATIVOS

El potencial de la utilización de datos genómicos en el entorno clínico y de investigación es considerable. No obstante, para acceder a las capacidades que nos ofrece esta tecnología debe existir un proceso activo de traslación que ayude a que el desarrollo de la investigación en este campo alcance una madurez adecuada para demostrar la utilidad clínica y el coste-efectividad de estas aproximaciones, identificando cómo pueden aplicarse en el ámbito clínico.

Para todo ello, deben adecuarse estructuras y procesos con el objetivo de facilitar su adopción y difusión, integrando nuevas prácticas en las vías ya establecidas que incluyan también ciertos requerimientos relativos a la formación de los profesionales implicados. Esto es especialmente relevante, puesto que la utilización de las tecnologías ómicas (tabla 1.1) puede implicar a todo el espectro de atención al paciente proporcionando un mayor conocimiento sobre el riesgo de enfermedad, mejorando la aplicación de acciones preventivas, y ayudando a que los pacientes reciban el tratamiento más eficaz en un menor plazo de tiempo.

Con el ritmo actual de desarrollo, es esperable que en los próximos 3 a 5 años puedan utilizarse muchas aplicaciones de la genómica actualmente en desarrollo. Esta visión es real y factible. No obstante, su desarrollo requiere que se tengan en cuenta una serie de puntos que incluyen:

- Procesos estandarizados y sólidos para establecer la validez clínica y la utilidad de las pruebas genómicas. Para asegurar la calidad de cada prueba en particular, tecnología utilizada y centro en que se realiza.
- Normas claras para el desarrollo e incorporación de pruebas genómicas dentro de las vías clínicas, proporcionando a los profesionales de la salud un proceso sencillo y universal en relación con la solicitud de pruebas y recepción de resultados.
- Una infraestructura bioinformática segura y robusta que permita obtener información útil de una forma rápida y con un coste adecuado en el entorno de variantes conocidas.
- Profesionales con las habilidades y conocimientos necesarios para hacer un uso efectivo de la tecnología genómica. Esto incluye llevar a cabo las pruebas, gestionar resultados de los datos y análisis, así como una mayor comprensión y conciencia de la genómica y su papel en el entorno clínico.
- Desarrollo del marco legal para abordar adecuadamente los complejos retos que la medicina genómica –y, en particular, la disponibilidad de la información genómica– crea, proporcionando así un entorno protegido para los datos genómicos.
- Un enfoque coordinado y coherente a nivel público para promover el entendimiento de la genómica y lo que significa para el cuidado de la salud.

Estos puntos pueden suponer un punto de partida para la integración de la genómica en la medicina convencional. Para ello, debe trabajarse de forma alineada en el ámbito clínico y de investigación.

Existen modelos de implantación que toman en cuenta estas consideraciones, como el propuesto en el Reino Unido³⁸, en el que se propone integrar la medicina genómica como parte de los procesos y estructuras existentes. Así, la recapitalización no debe ser la opción por defecto, teniendo en cuenta que no se obtendrá un servicio más efectivo por el hecho de que todos los centros cuenten con la última tecnología. Por el contrario, el primer reto que debe asumirse será identificar las maneras más eficaces de incorporar este tipo de metodologías en los programas y vías clínicas.

Cualquier infraestructura propuesta debe hacer frente no solo a la genómica actual, sino también a las demandas futuras, tales como nuevos avances tecnológicos, incremento en el número de pruebas disponibles o traslación al ámbito clínico a diferentes niveles asistenciales. Es previsible que el continuo desarrollo de las tecnologías ómicas mejore la precisión, velocidad y coste de las pruebas, proporcionando además un flujo constante de nuevos datos sobre los que se puedan tomar acciones por parte de todas las ramas de la medicina.

Un aspecto crucial en este punto es considerar que la infraestructura debe diseñarse basándose en la obtención de los resultados esperados, teniendo en cuenta la equidad en el acceso a las mismas y sin someter a una carga excesiva al sistema de financiación.

Tal como se ha detallado en el desarrollo del presente documento, la realización de estudios genéticos implica no solo ofrecer la determinación en sí, sino también el asesoramiento y consejo genético asociado al mismo. Para ello, en el modelo citado anteriormente se propone su realización en centros de referencia cuya distribución dependa de la población asignada a los mismos, considerándose la existencia de un centro cada 2-5 millones de personas.

Estos centros deberían trabajar en estrecha cooperación con entidades de ámbito tecnológico que actuarían como puente entre la investigación traslacional y la innovación en los centros de referencia, aportando el capital académico y científico necesario para asegurar la promoción, adopción y expansión del nuevo conocimiento obtenido. De manera complementaria, la incorporación de nuevas pruebas genéticas debería someterse a aprobación por comisiones creadas al efecto, con objeto de asegurar su calidad y validez clínica.

El rápido progreso de la genómica producido en los últimos años está también ligado al desarrollo de las tecnologías de la información. La capacidad de procesamiento de los ordenadores actuales hace posible el análisis de datos genómicos a una velocidad (y coste) abordable para la asistencia sanitaria. Paralelamente, la evolución en el almacenamiento de datos habilita el manejo y administración de las grandes cantidades de información generadas en estos procesos.

Hasta la fecha, el registro e interpretación de variantes genéticas con significado clínico se ha llevado a cabo utilizando bases de datos enfocadas a genes individuales, mantenidas por diferentes grupos con diferentes intereses, lo que ha supuesto ventajas como proporcionar recursos valiosos en términos de interpretación de variantes, e inconvenientes, como la falta de uniformidad en las interfaces utilizadas, soporte heterogéneo, accesibilidad a los datos o calidad de los mismos. Para resolver estas dificultades, el estudio Deciphering Developmental Disorders (DDD)³⁹ proporciona un modelo interesante de uso e intercambio de datos, que reúne a 23 centros regionales de genética del Reino Unido en la búsqueda de causas genéticas de síndromes malformativos de la infancia combinando recursos, como los microarrays y secuenciación masiva y experiencia en el análisis de resultados, con el objetivo de proporcionar un diagnóstico específico en el momento actual y mejorar para la realización de diagnósticos futuros. Otro ejemplo destacable del esfuerzo colaborativo de diferentes centros es el impulsado por el International Cancer Genome Consortium (ICGC)⁴⁰, donde se han propuesto como objetivo el caracterizar las anomalías genómicas de 50 tipos diferentes de cáncer. Para ello, han creado el ICGC Data Portal, una herramienta que permite a cada miembro del ICGC gestionar y mantener sus bases de datos de manera local a la vez que proporcionarles un acceso único a todos los datos genómicos generados por el consorcio⁴¹.

En este punto, la formación de los profesionales tiene un papel fundamental en la toma de conciencia y comprensión de las tecnologías genómicas. La incorporación más amplia de la perspectiva genómica de la patología dará lugar a un aumento considerable de la demanda de las pruebas genéticas, interpretación y gestión de datos genómicos. Por ello, deben articularse mecanismos de formación reglada lo que, en nuestro país, será probablemente una

consecuencia de la incorporación de la especialidad de Genética al Sistema Nacional de Salud. Esto se unirá al desarrollo de nuevos programas de máster, programas específicos en genética clínica, epidemiología o bioinformática.

Finalmente, ya se ha indicado extensamente en el presente documento la relación de la obtención y manejo de datos genómicos con importantes aspectos éticos y legales. Aunque no sea algo nuevo, sí que supone un cambio importante en la escala, lo que implica que es esencial desarrollar normas específicas aplicables a la medicina genómica dentro de un marco ético y legal que maximice el potencial de esta tecnología, minimizando los daños potenciales como el uso inadecuado, la estigmatización y la discriminación. Organizaciones como Wellcome Trust y la Human Genetics Commission del Reino Unido están actualmente desarrollando normas en relación con la utilización de consentimientos genéricos para este tipo de pruebas. También la Joint Committee on Medical Genetics ha generado un documento en relación con el uso compartido de datos genéticos⁴².

NECESIDADES DE PERSONAL Y FORMACIÓN

Para discutir las cuestiones relativas a la organización, planificación y sostenibilidad y los desafíos de las nuevas tecnologías genéticas, la European Society of Human Genetics (ESHG) celebró un simposio satélite sobre el “Cambio de paisaje de las pruebas genéticas”. La audiencia consistió eminentemente en una mezcla de genetistas de laboratorio, genetistas clínicos, especialistas en ética, sociólogos, representantes de organizaciones de pacientes y abogados. Se celebraron multitud de talleres y debates en grupos pequeños y se llegaron a conclusiones y acuerdos importantes, sobre todo enfocados a cuáles debían ser los posibles caminos para abordar correctamente estos cambios. Estaban sobre todo enfocados a mejorar y adaptar los servicios de genética para que los pacientes puedan recibir información precisa y pertinente, y que la gestión de la información clínica, de las bases de datos y de la información genómica, las conductas éticas y los consentimientos y los límites difusos entre la asistencia sanitaria y la investigación pudieran estar mejor definidos. También se ahondó en la gestión de la información en lo referente a los hallazgos inesperados o no solicitados (“hallazgos incidentales”). En el contexto del diagnóstico, las pruebas estándar para determinados problemas clínicos pueden producir resultados que no fueron buscados. En la discusión de las implicaciones de este fenómeno es importante tener en cuenta la distinción entre el análisis dirigido con paneles y el abordaje global con estudios de exomas o genomas.

Es por este motivo que la necesidad de personal cualificado y entrenado en todos los pasos del proceso es imprescindible para la correcta gestión del mismo. La implantación práctica del proceso requiere de actividad clínica de consulta, actividad de diagnóstico de laboratorio y un análisis bioinformático complejo y exhaustivo.

En el ámbito clínico, se requiere un asesoramiento previo y posterior al estudio genético proporcionado por personal entrenado en asesoramiento genético, con capacitación orientada al proceso de comunicación y con formación sólida en genética clínica y molecular. Este perfil es imprescindible para la correcta comunicación al paciente sobre la utilidad clínica del test que se va a ofrecer/realizar. Actualmente, cualquier evaluación fiable de la utilidad clínica de las nuevas tecnologías para las pruebas genéticas es obligada. Esta actividad debe ser ofertada en

servicios de genética hospitalarios y las recomendaciones actuales sugieren que debe evitarse por todos los medios que se hagan en atención primaria.

Dentro de los servicios de genética, el diagnóstico genético asistencial está empezando a utilizar la NGS (*Next Generation Sequencing*) como rutina y esto implica que se generan un sinnúmero de datos/información que no necesariamente está relacionado con la indicación médica del estudio, pero que podría usarse con fines asistenciales (de detección, diagnóstico y medicina personalizada). Los equipos médicos de genetistas clínicos están comenzando a discutir los resultados y las implicaciones más amplias de los resultados para el paciente y su familia y a dar consejos de salud basados en los datos generados de secuenciación de exomas o genomas. Esto traerá un cambio sustancial en el alcance de los genetistas clínicos y necesitará programas de capacitación para profesionales y público.

En el ámbito del laboratorio, los retos de la NGS desde esta perspectiva son la realización de las técnicas de extracción y preparación del ADN, la captura y la preparación de librerías para luego posteriormente proceder a su secuenciación masiva..

En el ámbito bioinformático se realiza el control de calidad de la muestra para poder determinar si la muestra es analizable o no, el alineamiento con el genoma de referencia y el posterior análisis para la identificación de “variantes normales” que carecen de importancia clínica o de variantes patológicas asociadas a problemas médicos. Para ello es necesario que todos los laboratorios de diagnóstico genético cuenten con personal entrenado y muy bien capacitado en la interpretación del significado clínico de la información obtenida y la capacidad de eventualmente validar estos resultados con otros métodos o técnicas.

Mientras que la biología y la biomedicina se transforman rápidamente en una ciencia de datos (*data science*) es fácil observar la falta generalizada de formación de muchos de los profesionales en estas áreas para afrontar esta situación. Es paradójico que en otras disciplinas científicas la mayor parte de los profesionales gestionan y analizan sus datos (pensemos en astrónomos, economistas o físicos nucleares), mientras que muy pocos biólogos son capaces de hacerlo. Esta situación crea una presión extraordinaria sobre los bioinformáticos que tienen que desarrollar su propia actividad científica creando nuevos algoritmos y métodos, y además deben hacer que los métodos sean fácilmente accesibles (trivialmente en muchos casos) a una comunidad de usuarios con muy poca o nula formación para su uso. Mientras que existe un esfuerzo directo en muchas universidades en países de nuestro entorno para corregir estos problemas, tanto para la formación en bioinformática, junto a principios básicos de matemáticas, física y estadística, para biólogos, como formación especializada para bioinformáticos (incluyendo la adopción de temarios acordados a nivel internacional), es notorio el retraso que las universidades españolas tienen en la formación de biólogos y otras áreas biomédicas en aspectos de bioinformática. Si el problema es grave en el caso de los biólogos y otras áreas biomédicas, la situación entre los profesionales en medicina parece aún más compleja, en parte por sus carencias adicionales en genética y genómica y en parte por la estricta organización en especialidades que deja poco lugar para la interpretación de datos genómicos transversales. Por otro lado, también se detecta una falta de formación por parte de los bioinformáticos de los problemas médicos que abordan en estos estudios. Se necesita por tanto planificar este tipo de formación para los profesionales que se dediquen a los análisis bioinformáticos en el ámbito de la biomedicina.

INFRAESTRUCTURAS

Las infraestructuras necesarias para la implementación de la NGS en asistencia hospitalaria dependerán en gran medida del volumen y complejidad de las patologías a asistir. Se considera recomendable la implementación actual de diagnóstico asistencial mediante NGS en hospitales terciarios (generalmente asociados a una universidad), para garantizar un volumen de pacientes y muestras que justifiquen la inversión y que mantengan la formación continua y la capacitación del personal que participa.

En relación con el material inventariable o aparatos, se requiere la infraestructura clásica de los laboratorios de genética, tales como termocicladores, secuenciadores Sanger clásicos del tipo de electroforesis capilar, y cada vez más, de la capacidad de validación con experimentos funcionales de algunas variantes. Esto se puede obtener con infraestructuras para validación *in vitro* de variantes (ensayos funcionales con cultivos celulares) o con validaciones más complejas o largas como modelos de validación *in vivo* con modelos animales (murinos o cada vez más pez cebra debido a su fácil manipulación, mantenimiento y tiempo de obtención de resultados). Estas nuevas herramientas están comenzando a implantarse lentamente en los hospitales como una forma más de validar algunas variantes de importancia médica.

En el ámbito bioinformático se requieren ordenadores específicos para poder ejecutar programas en diferentes lenguajes de programación, con altas capacidades de computación y de *clusters* o unidades de almacenamiento de ficheros informáticos de alta capacidad, ya que deben estar resguardados desde el punto de vista legal por varios años, y los ficheros que se utilizan habitualmente para análisis (FASTQ, BAM) son muy “pesados” desde el punto de vista informático (v. cap. 5).

Estas instalaciones deben garantizar el mantenimiento de los estándares de calidad. Los mecanismos deben ser instalados para excluir mezcla de muestras. Para cumplir esta función, existe la necesidad de contar con bases de datos de pacientes de libre disponibilidad/datos completos sobre todas las mutaciones conocidas y documentación detallada de las asociaciones con el problema clínico del paciente. Todo esto implica que la dotación en infraestructuras debe ser cuidadosamente planificada desde el inicio y la dotación en espacios y aparatos garantizada en el tiempo.

La interacción continua, preferiblemente en el mismo sitio físico de todos los actores (genetistas clínicos, genetistas de laboratorio e informáticos y bioinformáticos) son cruciales para el correcto desarrollo de todo el proceso.

El objetivo de la investigación es la generación de nuevos datos y conocimientos, y el de la atención clínica el de responder a las necesidades e intereses de los pacientes. Aunque en teoría esta distinción puede parecer clara, en la práctica la división entre ambos contextos es de hecho muy ambigua. Por ejemplo, no siempre es fácil para los pacientes distinguir las pruebas genéticas que se ofrecen como parte de su manejo clínico de las pruebas que se ofrecen como parte de un protocolo de investigación. Los datos de estudios también muestran que algunos investigadores son susceptibles de creer y entender que sus propias actividades de investigación son también una forma de proporcionar cuidado indirecto de los pacientes. Por último, las pruebas genéticas masivas actuales desdibujan muchas fron-

teras: entre los “consumidores” de pruebas genéticas y los verdaderos pacientes, así como entre los “consumidores” y los sujetos de investigación. Algunas investigaciones recientes han puesto de relieve las diversas preocupaciones que acompañan a la ausencia de fronteras claras entre la investigación y las pruebas genéticas asistenciales, sobre todo cuando la invitación a participar en una investigación está siendo solicitada por un profesional del sistema salud que está tratando a un paciente.

En resumen, como ocurre con frecuencia en la implementación de nuevas tecnologías y muy específicamente en la implementación de pruebas genéticas, las fronteras entre la investigación y los servicios asistenciales no es siempre sencillo de definir. Por ello, debido al coste y complejidad de infraestructuras para la genética (estructuras de laboratorio, aparatos, tecnologías y fungibles), es muy frecuente que los usos de los mismos sean compartidos, y tanto el personal como los aparatos u otras estructuras no se dupliquen y por lo tanto se optimicen con un uso compartido tanto para asistencia como para investigación. No obstante, es importante con el objetivo de la correcta gestión de los recursos, que pueda discernirse sobre cada muestra si es asistencial o de investigación, para poder eventualmente repercutir costes a un área u otra. Por otra parte, para los proyectos de investigación, se requieren formularios de consentimiento más específicos y un acuerdo claro sobre el uso de datos y almacenamiento de estos (anónima o codificada, etc.), que es totalmente distinto al proceso asistencial, donde el paciente debe estar totalmente identificado desde el inicio del mismo.

En otro orden de cosas, deben definirse en las instituciones sanitarias y los sistemas de atención sanitaria quiénes son los profesionales que pueden solicitar pruebas concretas. Esta atribución de funciones debe estar basada en la formación y competencia del profesional y en la complejidad de la propia prueba y las implicaciones de la información derivada del resultado.

Existen muy diversas situaciones en los centros de nuestro país para abordar esto:

- La ausencia de restricción en la solicitud de pruebas o su autorización basada solo en criterios económicos.
- La autorización previa por un ente superior (dirección del laboratorio, gerencia) basada en criterios médicos y económicos. La valoración de estos criterios es con frecuencia difícil para las personas o entidades que conceden la autorización, no conocedores de las pruebas y sus implicaciones.
- La restricción de la indicación de las pruebas a profesionales concretos y situaciones concretas.

La institución sanitaria se hace responsable del uso adecuado de la prueba, tanto en su solicitud como en la entrega de resultados. Los sistemas actuales no garantizan el uso óptimo de esta responsabilidad.

La reciente incorporación del asesoramiento o consejo genético en la cartera común básica de prestaciones del Sistema Nacional de la Salud (Orden SSI/2065/2014, de 31 de octubre, por la que se modifican los anexos I, II y III del Real Decreto 1030/2006, de 15 de septiembre, por el que se establece la cartera de servicios comunes del Sistema Nacional de Salud y el procedimiento para su actualización) destaca la necesidad de mejorar la formación de los profesionales sanitarios que deberán llevar a cabo el asesoramiento genético.

Los nuevos retos del asesoramiento genético en la era de la genómica y posgenómica son múltiples y diversos⁴³. Entre ellos destacan, además de mejorar la formación en genética y genómica del equipo profesional multidisciplinario que lleve a cabo el consejo, la necesidad de:

- a) ofrecer una información basada en la evidencia sobre la utilidad clínica de las variantes genómicas;
- b) entender y transmitir al paciente que estas técnicas pueden dar lugar al hallazgo de información inesperada no buscada con implicaciones clínicas, así como información con consecuencias aún inciertas; estos hallazgos son mucho más numerosos que con otros análisis previamente disponibles, y más en estos momentos de desarrollo de un conocimiento aún muy incompleto;
- c) adaptar los consentimientos informados;
- d) desarrollar nuevos modelos clínicos para mejorar el acceso y la comunicación responsable de la información compleja sobre riesgo de enfermar;
- e) contribuir al desarrollo y funcionamiento de programas de prevención y supervisión de salud adaptados a las variantes genómicas del individuo, así como contribuir a la investigación que permita nueva evidencia sobre su significado, y
- f) conocer los peligros potenciales que plantean la prematura comercialización de la primera generación de perfiles genómicos.

Tal y como indica Kennet Offit, “la genómica personalizada, en lugar de constituir un nuevo paradigma, debe considerarse como un avance a partir del modelo construido durante varias décadas en la práctica de la medicina genética”⁴⁴.

IMPACTO ECONÓMICO EN ASISTENCIA E INVESTIGACIÓN

Al evaluar el impacto de la implementación NGS en una organización, un parámetro clásico de comparación frecuentemente citado es el coste de secuenciar un número de pares de bases. Durante años, los institutos nacionales de investigación, National Human Genome Research Institute (NHGRI) ha rastreado sus costes para secuenciar un millón de pares de bases (que se refiere como el coste de la secuenciación por cada megabase, NHGRI, 2010). Los números del NHGRI muestran una curva denominada ley de Moore desde el inicio del año 2000 hasta alrededor de principios de 2008. Los costes variaron de alrededor de 10 000 dólares por megabase a mediados de 2001 a menos de 1000 dólares a principios de 2008. En 2008, los entonces llamados sistemas de secuenciación NGS llegaron al mercado y sus mejoras tecnológicas aceleraron la reducción de precio a un ritmo que superó a la ley de Moore. Desde la introducción de los sistemas de NGS, el costo por megabase se ha reducido a la mitad más o menos cada 5 meses. A finales de 2011, el NHGRI informó que el coste es de más o menos 0,10 dólares por megabase para la secuenciación.

Si bien estas cifras y otras similares de otros estudios son siempre muy citadas, lo que a menudo no se menciona o se sabe es lo que esos costes incluyen. Esos costes menos conocidos o costes ocultos pueden tener un impacto significativo en el presupuesto de una organización.

Por ejemplo, los números del NHGRI incluyen el coste de mano de obra de preparación de muestras y para los fungibles. Sin embargo, el número no da una indicación real del coste

total por estos conceptos; el NHGRI observó que los costes de los consumibles para secuenciadores actuales varían mucho según la marca y el tipo de secuenciador y pueden sumar hasta 1 millón de dólares adicional o de diferencia por año/máquina. Los cálculos del NHGRI también incluyen una tasa de amortización para el *hardware* de 5 años, mientras que con el cambio y recambio de secuenciadores debido a que la tecnología se queda obsoleta, una proyección a 3 años es ahora una amortización más realista.

A pesar de su reciente desarrollo, el coste de la secuenciación mediante técnicas de NGS ha experimentado un notable descenso en los últimos años. En líneas generales, puede afirmarse que las técnicas de NGS permiten determinar en la actualidad la secuencia de fragmentos de ADN de una manera más rápida y económica que la secuenciación tradicional. Así, al menos en grandes proyectos, el coste medio estimado se sitúa en torno a los 0,5 dólares por kilobase (kb) de ADN estudiado mediante secuenciación capilar, mientras que tendría un coste similar pero por cada megabase estudiada en el caso de la NGS, lo que supone un coste aproximadamente mil veces menor⁴⁵. **Un ejemplo significativo es que el coste de secuenciar un solo gen de la manera tradicional es el mismo que el coste de secuenciar todos los genes humanos mediante las técnicas NGS.** En la figura 6.1 puede apreciarse de forma gráfica el descenso exponencial experimentado en los últimos años en el coste de las técnicas de secuenciación tras la introducción de la NGS, si bien, parece existir una tendencia hacia la estabilización en los dos últimos años.

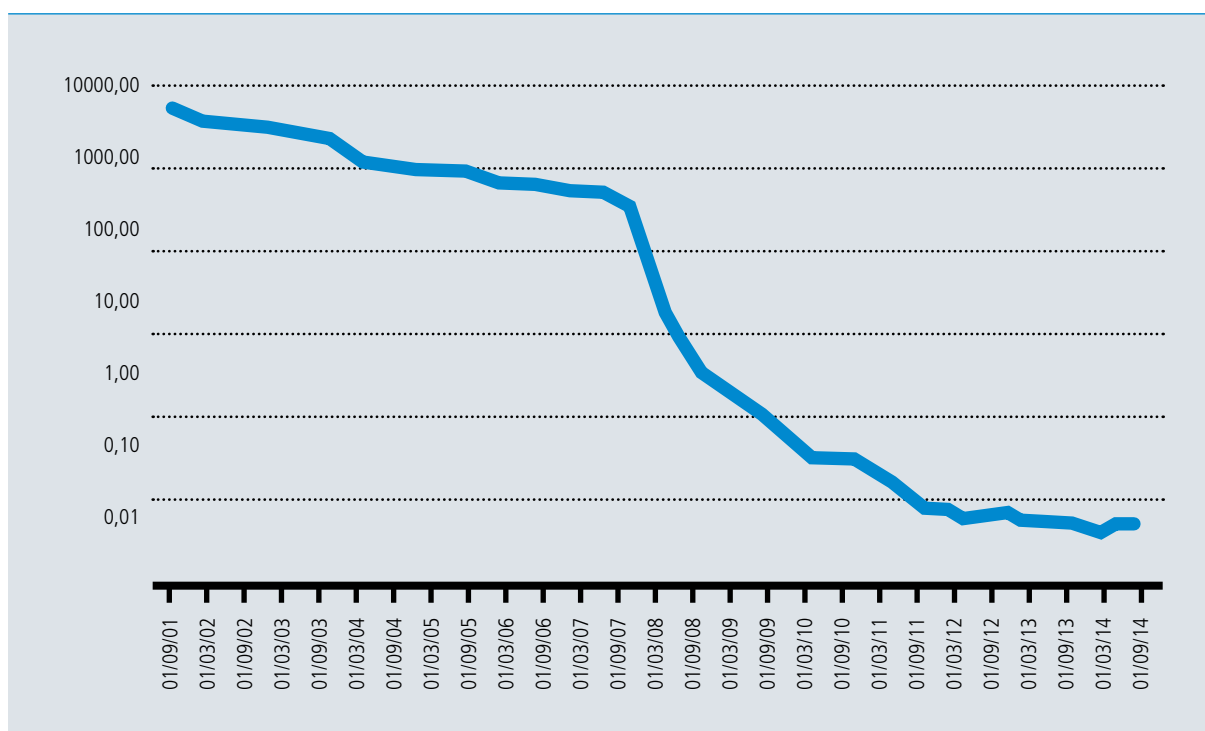


Figura 6.1. Evolución del coste de secuenciación por megabase de ADN en dólares.

Fuente: National Human Genome Research Institute (NHGRI)⁴⁶.

Asimismo, el desarrollo tecnológico ha permitido a las compañías del sector ofrecer equipos de NGS con una capacidad de procesamiento y precio acordes con la práctica asistencial, lo que unido al desarrollo de protocolos comerciales estandarizados de secuenciación está facilitando su implantación en diversos hospitales y centros del Sistema Nacional de Salud. Sin embargo, en la práctica clínica o en proyectos de investigación de pequeña o media envergadura, el coste real de las técnicas de NGS dista mucho de haberse reducido en la misma proporción que en los grandes proyectos de secuenciación⁴⁷.

Así, y como se ha comentado a lo largo de todo este documento, los rápidos avances tecnológicos y la disminución de precios significativos en las tecnologías de NGS han permitido generar técnicas de investigación sofisticadas para las organizaciones, grupos y personas que antes no podían permitirse el equipo. Esto podría cambiar radicalmente la investigación en ciencias de la vida, lo que permite una investigación más exhaustiva sobre la contribución genética a las enfermedades y proporcionar un mejor conocimiento para el diseño y composición de nuevos fármacos candidatos, todo en tiempos más rápidos de lo que era posible antes. Por desgracia, hay varios obstáculos para explotar todo el potencial de la tecnología. Mientras que la NGS permite tener resultados más rápido y son menos costosos que los secuenciadores clásicos (secuenciadores Sanger), el tiempo y el costo para preparar muestras para la secuenciación NGS y analizar los resultados de la secuenciación puede obviar esos beneficios. Por ejemplo, la cantidad de datos que salen de los nuevos sistemas de NGS y los enfoques cada vez más multidisciplinarios para la investigación en ciencias de la vida hace que la interpretación de los datos sea mucho más compleja. Este tiempo de interpretación y sus costes asociados son fundamentales para la genómica clínica; aunque la riqueza de información es abrumadoramente mayor para estudios de NGS que con la secuenciación clásica. Esto desde luego es distinto si se trata de pruebas asistenciales o de investigación.

Si bien no es el propósito de este documento hacer un análisis pormenorizado de los costes de las técnicas de NGS, que vendrá condicionado en buena medida por la estrategia diagnóstica elegida y el equipo concreto empleado, sí conviene reseñar una serie de variables que deberán tenerse en cuenta a la hora de realizar un análisis de costes:

- Personal (técnicos, genetistas, bioinformáticos, etc.).
- Extracción, cuantificación y control de calidad de ácidos nucleicos.
- Reactivos y consumibles generales de laboratorio.
- Reactivos y consumibles relacionados con la secuenciación (preparación de librerías, generación del *template* o *cluster* de secuenciación, etc.).
- Costes de validación de las alteraciones candidatas.
- Costes derivados de la repetición de determinaciones por errores y/o confirmaciones.
- Costes derivados de garantizar la trazabilidad de las muestras y la confidencialidad de los resultados obtenidos.
- Amortización y mantenimiento de los equipos de secuenciación y equipos auxiliares (p. ej., termocicladores, espectrofotómetro, etc.).
- Amortización y mantenimiento del equipamiento informático y de almacenamiento de datos.
- Análisis bioinformático de los datos (ensamblaje de las secuencias, alineamiento con genoma de referencia, anotación de variantes, eventual desarrollo de herramientas bioinformáticas, etc.).

- Interpretación de los resultados. Eventual acceso a bases de datos genómicas de pago asociadas a patología humana.
- Controles de calidad internos y externos. Gastos de certificación y/o acreditación. Formación continuada del personal.
- Costes indirectos, tanto relacionados con el laboratorio de secuenciación (p. ej., electricidad, climatización, telefonía, etc.) como con la gestión de la propia organización (p.ej., gestión de residuos, contabilidad, suministros, etc.).

Los gastos de NGS pueden encuadrarse dentro de dos grandes categorías: los gastos de capital (CAPEX, *capital expenses*) y los gastos operativos (OPEX, *operative expenses*). Por el lado de CAPEX, todos los sistemas de NGS han bajado de precio. Pero las organizaciones de ciencias de la vida deben considerar los costes OPEX que incluyen los gastos para preparar muestras, analizar los datos que emiten las máquinas, y gestionar el almacenamiento y el acceso a esos datos en sistemas informáticos fiables.

Además, el coste por megabase no incluye los gastos de análisis e interpretación. Estos costes deben incluir dinero para comprar servidores más potentes y estaciones de trabajo, *software* de sistemas operativos y más licencias para el análisis de los usuarios finales (los que van a emitir el informe clínico asistencial de laboratorio). Por último, el personal y los gastos generales para interpretar datos podrían sesgar los costes e incrementar mucho los OPEX, ya que las necesidades de horas por profesional que requieren los análisis de estudios de NGS son bien distintos a los que requería la secuenciación capilar o Sanger; una vez más, difícilmente comparables por la diferente riqueza y profundidad de datos que dan los resultados de NGS en comparación con la secuenciación clásica. Cabe señalar que a pesar de que la gestión de datos y de ficheros informáticos bajarán algo los precios y consecuentemente los costes operativos, la demanda de almacenamiento seguirá aumentando debido al incremento en el número total de secuencias producidas y el número total de análisis que se realizan.

En la mayoría de los cálculos de costes de los organismos de investigación, el coste por megabase no incluye el almacenamiento de datos y los costes de gestión, aunque algunos centros sí los consideran. La disminución de los costes de NGS significa que más laboratorios están ejecutando más secuencias, produciendo más y más datos o que se están usando secuenciadores más potentes que en cierta medida abaratan el coste por megabase. Un documento (Stein, 2010) señala que el coste por megabyte de almacenamiento bruto se estaba reduciendo a la mitad cada 14 meses durante la última década. Sin embargo, la simple adición de capacidad de almacenamiento informático puede aumentar los costos, ya que los dispositivos de almacenamiento adicionales necesitarán ser gestionados, requerirán más espacio en el rack, y precisarán más electricidad para energía y refrigeración. Por otra parte, la naturaleza cada vez más multidisciplinar de la investigación en ciencias de la vida está afectando a los flujos de trabajo de computación y análisis aumentando exponencialmente la demanda de rendimiento en los sistemas de almacenamiento. Los sistemas de NGS de hoy producen conjuntos de datos mucho más ricos, por lo que los datos son útiles para los investigadores de diferentes disciplinas. Por ejemplo, es bastante común que los científicos especializados en la bioinformática, genómica, proteómica, biología de sistemas, y otras áreas de investigación y clínicos compartan y utilicen los mismos datos experimentales. En diferentes momentos del ciclo de I+D, cada disciplina utiliza sus propias herramientas de análisis, cada uno con diferentes rendimientos y requisitos. Esto lleva a que las cargas de trabajo sean altamente impredecibles en los sistemas

de almacenamiento. Tomando estos factores de infraestructura en cuenta, lo que se necesita para apoyar la investigación en ciencias de la vida de hoy es una solución de almacenamiento que sea altamente escalable en capacidad y rendimiento.

Adquisición de equipos de NGS

Una de las principales decisiones que deben adoptarse a la hora de integrar las técnicas de NGS en las unidades de diagnóstico genético es la adquisición de los equipos de secuenciación. Ello conlleva inclinarse por un tipo de tecnología en particular; en el momento actual está muy polarizado en dos de ellas: la secuenciación basada en sistemas ópticos de detección y la secuenciación basada en semiconductores. Aunque ambas tecnologías tienen conocidas ventajas e inconvenientes, en líneas generales sus resultados son comparables. Respecto a la capacidad de los equipos, la mayoría de los laboratorios clínicos optan por la instalación de equipos de sobremesa con un precio más ajustado y con una capacidad de secuenciación más que suficiente para el empleo de paneles de resecuenciación dirigida de pequeño-mediano tamaño. El uso de equipos con capacidad para secuenciar exomas o genomas completos está, en general, limitado a grandes centros de investigación, no tanto por las dificultades técnicas, sino por la complejidad del análisis posterior.

Al margen de las características técnicas, entre los aspectos más importantes a la hora de inclinarse por la adquisición de una plataforma de secuenciación u otra destacan la duración de la garantía, el servicio posventa o el coste de los contratos de mantenimiento de los equipos (invariablemente desproporcionados). En este sentido, conviene tener presente que la adquisición de estos equipos supone un desembolso importante para los centros públicos con la certeza de que quedarán obsoletos en un plazo relativamente corto de tiempo. En este escenario, otras alternativas tales como el alquiler o la “cesión por consumo”, que obvian los elevados gastos de mantenimiento, pueden ser consideradas.

Estrategias de secuenciación y su impacto en el coste

En el ámbito de la genética y de la biología en general, se distinguen dos tipos de secuenciación: la secuenciación *de novo*, que consiste en la obtención de la secuencia de un genoma desconocido hasta ese momento (p. ej., una bacteria, un animal o una planta) y la resecuenciación o secuenciación de un genoma para compararlo con otro genoma de la misma especie que se considera consenso.

La secuenciación llevada a cabo con fines clínicos, independientemente de la estrategia que se utilice, es una resecuenciación, ya que compara el ADN de un paciente o grupo frente a un genoma humano de referencia que está en constante actualización (actualmente se trabaja con la versión 38).

- **Resecuenciación dirigida:** en la práctica clínica consiste habitualmente en la secuenciación de uno o varios genes agrupados en un “panel” y cuya relación con un trastorno concreto está bien definida. Entre las ventajas económicas de este tipo de aproximación figuran la facilidad en el manejo de datos y su interpretación, así como el menor riesgo de hallazgos secundarios. Además, existen en el mercado un número creciente de paneles pre-diseñados optimizados para diversas patologías, algunos de los cuales están validados para

diagnóstico clínico. A menudo estos productos se acompañan de algoritmos de *software* que simplifican enormemente el análisis y que lo hacen accesible a laboratorios sin experiencia previa en complejos análisis bioinformáticos. Además, existe la posibilidad de diseñar, de una forma relativamente sencilla, paneles a medida o “custom” optimizados según las necesidades de cada laboratorio. El diseño de estos paneles debe ser especialmente cuidadoso, ya que suponen un desembolso inicial importante y, a menudo, la repercusión en el coste real por muestra resulta superior a la prevista, considerando que, a excepción de varios tipos de cáncer o leucemias, la mayoría de enfermedades genéticas son de baja prevalencia y, por tanto, el número de casos susceptibles de ser secuenciados en un mismo panel suele ser pequeño. Por otra parte, un diseño inadecuado o la falta de previsión en una patología sujeta a nuevos descubrimientos puede obligar al rediseño de estos con la consiguiente repercusión en el coste.

Paradójicamente, la mayor limitación económica de este tipo de tecnología deriva de la enorme capacidad de secuenciación de los equipos actuales, incluso aquellos pensados para el diagnóstico clínico. Así, para ser realmente efectiva en costes, es necesario multiplexar muestras y secuenciarlas en un mismo experimento. En el caso de una enfermedad relativamente rara, esta puede no ser una opción válida o suponer retrasos prolongados antes de poder disponer de un número suficiente de muestras para realizar un diagnóstico molecular. Del mismo modo, impone una limitación al tamaño mínimo de los laboratorios de diagnóstico, ya que solo aquellos con una casuística suficiente (normalmente laboratorios de referencia) pueden reducir de forma significativa el coste de las determinaciones.

Entre los aspectos técnicos clave a tener en cuenta tanto para la fiabilidad del diagnóstico molecular como por su repercusión en el coste de la secuenciación destaca la profundidad de lectura media o profundidad de cobertura. En general, a mayor profundidad de lectura mayor será la calidad de las secuencias obtenidas y menor el número de muestras que pueden incluirse en un mismo experimento. Encontrar un equilibrio entre ambas variables permitirá optimizar el coste de una determinación. El estudio de mutaciones en línea germinal requiere, en líneas generales, de una menor profundidad de lectura (p. ej., 100-500x), mientras que para la correcta identificación de mutaciones somáticas (fundamentalmente en cáncer) es deseable una mayor profundidad de lectura (p. ej., 2000x). Por otra parte, la cobertura de las regiones estudiadas es otro de los factores que hay que tener en cuenta en la medida que puede obligar, en determinados casos, a realizar determinaciones adicionales mediante otras técnicas para cubrir las regiones que no se secuencian adecuadamente mediante NGS.

RESUMEN DEL CAPÍTULO

Las tecnologías de alto rendimiento están cambiando rápidamente el campo de la detección de enfermedades genéticas. Por ello, su aplicación responsable tiene un gran impacto en el Sistema Nacional de Salud, así como en la formación y en la preparación de los trabajadores de la salud para decidir y discutir cuándo estas tecnologías son o no son adecuadas para determinadas indicaciones médicas en el plano de la medicina asistencial y en investigación biomédica. En este capítulo se aborda la organización, planificación y sostenibilidad de estas nuevas necesidades, así como también las necesidades de personal y formación y su impacto económico, tanto en el ámbito hospitalario como en infraestructuras de investigación.

Existen varios modelos organizativos de cómo abordar esta situación. Para acceder a las capacidades que nos ofrece esta tecnología debe existir un proceso activo de traslación que ayude a que el desarrollo de la investigación en este campo alcance una madurez adecuada para demostrar la utilidad clínica y el coste-efectividad de estas aproximaciones, identificando cómo se pueden aplicar en el ámbito clínico.

El desarrollo e implementación de la medicina genómica requiere que se tengan en cuenta una serie de puntos que incluyen:

- a) procesos estandarizados y sólidos para establecer la validez clínica y la utilidad de las pruebas genómicas, para asegurar la calidad de cada prueba en particular, tecnología utilizada y centro en que se realiza;
- b) normas claras para el desarrollo e incorporación de pruebas genómicas dentro de las vías clínicas, proporcionando a los profesionales de la salud un proceso sencillo y universal en relación con la solicitud de pruebas y recepción de resultados;
- c) una infraestructura bioinformática segura y robusta que permita obtener información útil de una forma rápida y con un coste adecuado en el entorno de variantes conocidas;
- d) profesionales con las habilidades y conocimientos necesarios para hacer un uso efectivo de la tecnología genómica;
- e) desarrollo del marco legal para abordar adecuadamente los complejos retos que la medicina genómica –y, en particular, la disponibilidad de la información genómica– crea, proporcionando así un entorno protegido para los datos genómicos, y
- f) un enfoque coordinado y coherente a nivel público para promover el entendimiento de la genómica y lo que significa para el cuidado de la salud.

Existen varios modelos de implantación que toman en cuenta estos puntos en el Reino Unido, Holanda, etc. En este punto, la formación de los profesionales tiene un papel fundamental en la toma de conciencia y comprensión de las tecnologías genómicas. Es por este motivo que la necesidad de personal cualificado y entrenado en todos los pasos del proceso es imprescindible para la correcta gestión del mismo. A nivel práctico, el proceso requiere de al menos tres procesos separados e interrelacionados:

- a) actividad clínica de consulta y asesoramiento,
- b) actividad de diagnóstico de laboratorio y
- c) análisis bioinformáticos complejos y exhaustivos.

Las infraestructuras necesarias para la implementación de la NGS en asistencia hospitalaria dependerán en gran medida del volumen y complejidad de las patologías a asistir. Se considera recomendable la implementación actual de diagnóstico asistencial mediante NGS en hospitales terciarios (generalmente asociados a una universidad), para garantizar un volumen de pacientes y muestras que justifiquen la inversión y que mantengan la formación continua y capacitación del personal que participa.

En relación con inventariables o aparatos, se requiere la estructura básica de los laboratorios de genética, además de los secuenciadores masivos que se utilicen. En el ámbito bioinformático se requieren ordenadores específicos para poder ejecutar programas en diversos lenguajes de programación, con altas capacidades de computación y de *clusters* o unidades de almacenamiento de ficheros informáticos de alta capacidad, ya que deben estar resguardados desde el punto de vista legal por varios años, y los ficheros que se utilizan habitualmente para análisis (FASTQ, BAM) son muy “pesados” desde el punto de vista informático. Estas instalaciones deben garantizar el mantenimiento de los estándares de calidad. La interacción continua, preferiblemente en el mismo sitio físico de todos los actores (genetistas clínicos, genetistas y especialistas de laboratorio, informáticos y bioinformáticos) son cruciales para el correcto desarrollo de todo el proceso.

Deben definirse además en las instituciones sanitarias y los sistemas de atención sanitaria quiénes son los profesionales que pueden solicitar pruebas concretas. Esta atribución de funciones debe estar basada en la formación y competencia del profesional y en la complejidad de la propia prueba y las implicaciones de la información derivada del resultado. La reciente incorporación del asesoramiento o consejo genético en la cartera común básica de prestaciones del Sistema Nacional de la Salud (Orden SSI/2065/2014, de 31 de octubre, por la que se modifican los anexos I, II y III del Real Decreto 1030/2006, de 15 de septiembre, por el que se establece la cartera de servicios comunes del Sistema Nacional de Salud y el procedimiento para su actualización), destaca la necesidad de mejorar la formación de los profesionales sanitarios que deberán llevar el asesoramiento genético.

Al evaluar el impacto de la implementación de la NGS en una organización, un parámetro clásico de comparación frecuentemente citado es el coste de secuenciar un número de pares de bases. Desde la introducción de los sistemas de NGS, el costo por megabase se ha reducido a la mitad más o menos cada 5 meses. A finales de 2011, el NHGRI informó que el coste es de más o menos 0,10 dólares por Mb para la secuenciación. Así, a pesar de su reciente desarrollo, el coste de la secuenciación mediante técnicas de NGS ha experimentado un notable descenso en los últimos años. En líneas generales, puede afirmarse que las técnicas de NGS permiten determinar en la actualidad la secuencia de fragmentos de ADN de una manera más rápida y económica que la secuenciación tradicional.

Asimismo, el desarrollo tecnológico ha permitido a las compañías del sector ofrecer equipos de NGS con una capacidad de procesamiento y precio acordes con la práctica asistencial, lo que unido al desarrollo de protocolos comerciales estandarizados de secuenciación está facilitando su implantación en diversos hospitales y centros del Sistema Nacional de Salud. Las variables que deberán tenerse en cuenta a la hora de realizar un análisis de costes para la implementación de un programa que incluya tecnología de NGS son los costes de:

- a) personal (técnicos, personal especializado, bioinformáticos, etc.);
- b) extracción, cuantificación y control de calidad de ácidos nucleicos;

- c) reactivos y consumibles generales de laboratorio;
- d) reactivos y consumibles relacionados con la secuenciación (preparación de librerías, generación del *template* o *cluster* de secuenciación, etc.);
- e) costes de validación de las alteraciones candidatas;
- f) costes derivados de la repetición de determinaciones por errores y/o confirmaciones;
- g) costes derivados de garantizar la trazabilidad de las muestras y la confidencialidad de los resultados obtenidos;
- h) amortización y mantenimiento de los equipos de secuenciación y equipos auxiliares (p. ej., termocicladores, espectrofotómetro, etc.);
- i) amortización y mantenimiento del equipamiento informático y de almacenamiento de datos;
- j) análisis bioinformático de los datos (ensamblaje de las secuencias, alineamiento con genoma de referencia, anotación de variantes, eventual desarrollo de herramientas bioinformáticas, etc.);
- k) interpretación de los resultados; eventual acceso a bases de datos genómicas de pago asociadas a patología humana;
- l) controles de calidad internos y externos; gastos de certificación y/o acreditación, formación continuada del personal;
- m) costes indirectos, tanto relacionados con el laboratorio de secuenciación (p. ej., electricidad, climatización, telefonía, etc.) como con la gestión de la propia organización (p. ej., gestión de residuos, contabilidad, suministros, etc.). Los gastos, a su vez, de NGS pueden encuadrarse dentro de dos grandes categorías: los gastos de capital (CAPEX, *capital expenses*) y los gastos operativos (OPEX, *operative expenses*). Por el lado de CAPEX, todos los sistemas de NGS han bajado de precio. Pero las organizaciones de ciencias de la vida deben considerar los costes OPEX que incluyen los gastos para preparar muestras, analizar los datos que emiten las máquinas, y gestionar el almacenamiento y el acceso a esos datos en sistemas informáticos fiables.

Una de las principales decisiones que deben adoptarse a la hora de integrar las técnicas de NGS en las unidades de diagnóstico genético es la adquisición de los equipos de secuenciación. Al margen de las características técnicas, entre los aspectos más importantes a la hora de inclinarse por la adquisición de una plataforma de secuenciación u otra destacan la duración de la garantía, el servicio posventa o el coste de los contratos de mantenimiento de los equipos (invariablemente desproporcionados). En este sentido, conviene tener presente que la adquisición de estos equipos supone un desembolso importante para los centros públicos con la certeza de que quedarán obsoletos en un plazo relativamente corto de tiempo. En este escenario, otras alternativas tales como el alquiler o la “cesión por consumo”, que obvian los elevados gastos de mantenimiento, pueden ser consideradas.

1. Lander ES. Initial impact of the sequencing of the human genome [review]. *Nature*. 2011 Feb 10;470(7333):187-97. doi: 10.1038/nature09792.
2. Mardis ER. Next-generation sequencing platforms [review]. *Annu Rev Anal Chem (Palo Alto Calif)*. 2013;6:287-303. doi: 10.1146/annurev-anchem-062012-092628.
3. Lupski JR, Reid JG, Gonzaga-Jauregui C, Rio Deiros D, Chen DC, Nazareth L, et al. Whole-genome sequencing in a patient with Charcot-Marie-Tooth neuropathy. *N Engl J Med*. 2010 Apr 1;362(13):1181-91. doi: 10.1056/NEJMoa0908094.
4. Puente XS, Quesada V, Osorio FG, Cabanillas R, Cadiñanos J, Fraile JM, et al. Exome sequencing and functional analysis identifies BANF1 mutation as the cause of a hereditary progeroid syndrome. *Am J Hum Genet*. 2011 May 13;88(5):650-6. doi: 10.1016/j.ajhg.2011.04.010.
5. Yang Y, Muzny DM, Reid JG, Bainbridge MN, Willis A, Ward PA, et al. Clinical whole-exome sequencing for the diagnosis of mendelian disorders. *N Engl J Med*. 2013 Oct 17;369(16):1502-11. doi: 10.1056/NEJMoa1306555.
6. Puente XS, Pinyol M, Quesada V, Conde L, Ordóñez GR, Villamor N, et al. Whole-genome sequencing identifies recurrent mutations in chronic lymphocytic leukaemia. *Nature*. 2011 Jun 5;475(7354):101-5. doi: 10.1038/nature10113.
7. Cancer Genome Atlas Research Network. Comprehensive molecular characterization of gastric adenocarcinoma. *Nature*. 2014 Sep 11;513(7517):202-9. doi: 10.1038/nature13480.
8. Norton ME, Jacobsson B, Swamy GK, Laurent LC, Ranzini AC, Brar H, et al. Cell-free DNA analysis for noninvasive examination of trisomy. *N Engl J Med*. 2015 Apr 23;372(17):1589-97. doi: 10.1056/NEJMoa1407349.
9. Stratton MR, Campbell PJ, Futreal PA. The cancer genome [review]. *Nature*. 2009 Apr 9;458(7239):719-24. doi: 10.1038/nature07943.

10. Chang DK, Grimmond SM, Evans TR, Biankin AV. Mining the genomes of exceptional responders. *Nat Rev Cancer*. 2014 May;14(5):291-2.
11. American College of Medical Genetics (ACMG) 2014.
12. Loman NJ, Watson M. Successful test launch for nanopore sequencing. *Nat Methods*. 2015 Apr;12(4):303-4. doi: 10.1038/nmeth.3327.
13. Koch L. Technology: SMRT move? *Nat Rev Genet*. 2014 Mar;15(3):146. doi: 10.1038/nrg3678.
14. Xu X, Hou Y, Yin X, Bao L, Tang A, Song L, et al. Single-cell exome sequencing reveals single-nucleotide mutation characteristics of a kidney tumor. *Cell*. 2012 Mar 2;148(5):886-95. doi: 10.1016/j.cell.2012.02.025.
15. Wang Z, Gerstein M, Snyder M. RNA-Seq: a revolutionary tool for transcriptomics [review]. *Nat Rev Genet*. 2009 Jan;10(1):57-63. doi: 10.1038/nrg2484.
16. 1000 Genomes Project Consortium, Abecasis GR, Altshuler D, Auton A, Brooks LD, Durbin RM, Gibbs RA, et al. A map of human genome variation from population-scale sequencing. *Nature*. 2010 Oct 28;467(7319):1061-73. doi: 10.1038/nature09534. Erratum in: *Nature*. 2011 May 26;473(7348):544. Xue, Yali [added]; Cartwright, Reed A [added]; Altshuler, David L [corrected to Altshuler, David]; Keibel, Andrew [corrected to Keebler, Jonathan]; Koko-Gonzales, Paula [corrected to Kokko-Gonzales, Paula]; Nickerson, Debbie A [corrected to Nickerson, Deborah A].
17. Jones S, Anagnostou V, Lytle K, Parpart-Li S, Nesselbush M, Riley DR, et al. Personalized genomic analyses for cancer mutation discovery and interpretation. *Sci Transl Med*. 2015 Apr 15;7(283):283ra53. doi: 10.1126/scitranslmed.aaa7161.
18. Morgan XC, Huttenhower C. Meta'omic analytic techniques for studying the intestinal microbiome. *Gastroenterology*. 2014 May;146(6):1437-1448.e1. doi: 10.1053/j.gastro.2014.01.049. 19. Danecek P, Auton A, Abecasis G, Albers CA, Banks E, DePristo MA, et al; 1000 Genomes Project Analysis Group. The variant call format and VCFtools. *Bioinformatics*. 2011 Aug 1;27(15):2156-8. doi: 10.1093/bioinformatics/btr330.
20. Johnston JJ, Rubinstein WS, Facio FM, Ng D, Singh LN, Teer JK, et al. 1. Secondary variants in individuals undergoing exome sequencing: screening of 572 individuals identifies high-penetrance mutations in cancer-susceptibility genes. *Am J Hum Genet*. 2012 Jul 13;91(1):97-108.
21. Biesecker BB, Klein W, Lewis KL, Fisher TC, Wright MF, Biesecker LG, et al. How do research participants perceive "uncertainty" in genome sequencing? *Genet Med*. 2014 Dec;16(12):977-80.
22. Li MX, Gui HS, Kwan JS, Bao SY, Sham PC. A comprehensive framework for prioritizing variants in exome sequencing studies of Mendelian diseases. *Nucleic Acids Res*. 2012;40(7):e53.

23. Richards S, Aziz N, Bale S, Bick D, Das S, Gastier-Foster J, et al. Standards and guidelines for the interpretation of sequence variants: a joint consensus recommendation of the American College of Medical Genetics and Genomics and the Association for Molecular Pathology. *Genet Med*. 2015 Mar 5. doi: 10.1038/gim.2015.30.
24. MacArthur DG, Manolio TA, Dimmock DP, Rehm HL, Shendure J, Abecasis GR, et al. Guidelines for investigating causality of sequence variants in human disease. *Nature* 2014 April 24;508(7497):469-76.
25. Red Biobancos. Instituto de Salud Carlos III. Disponible en: <http://www.redbiobancos.es>
26. Wolf SM, Crock BN, Van Ness B, Lawrenz F, Kahn JP, Beskow LM, et al. Managing incidental findings and research results in genomic research involving biobanks and archived data sets. *Genet Med*. 2012 Apr;14(4):361-84. doi: 10.1038/gim.2012.23.
27. Bledsoe MJ, Clayton EW, McGuire AL, Grizzle WE, O'Rourke PP, Zeps N. Return of research results from genomic biobanks: cost matters. *Genet Med*. 2013 Feb;15(2):103-5. doi: 10.1038/gim.2012.105. Epub 2012 Aug 30.
28. Wolf SM. Return of results in genomic biobank research: ethics matters. *Genet Med*. 2013 Feb;15(2):157-9. doi: 10.1038/gim.2012.162.
29. Dove ES, Knoppers BM, Zawati MH. Towards and ethics safe harbour for global biomedical research. *J Law Biosci* 2014;1/1:3-51.
30. Nicolás P. Las bases nitrogenadas en bases de datos. El régimen jurídico de los archivos de información genética con fines de investigación biomédica. *Com Prop Ind Derch Comp*. 2013;68:191-224.
31. Guía para clientes que contraten servicios de Cloud Computing de la Agencia Española de Protección de Datos.
32. Guía de Seguridad de la Agencia Española de Protección de Datos. Disponible en: https://www.agpd.es/portalwebAGPD/canaldocumentacion/publicaciones/common/Guias/GUIA_SEGURIDAD_2010.pdf
33. Guía para los Miembros de los Comités de Ética de Investigación. Comité Director de la Bioética (CDBI), Consejo de Europa. 2011.
34. Alfonso I, Nicolás P, Fernández E. Aspectos éticos y legales relativos al tratamiento de muestras biológicas humanas en ensayos clínicos con medicamentos. Propuesta de criterios de evaluación para los comités éticos de investigación clínica (en prensa).
35. Ayuso C, Millán JM, Mancheño M, Dal-Ré R. Informed consent for whole-genome sequencing studies in the clinical setting. Proposed recommendations on essential content and process. *Eur J Hum Genet*. 2013 Oct;21(10):1054-9. doi: 10.1038/ejhg.2012.297.

36. Anticipate and Communicate. Ethical Management of Incidental and Secondary Findings in the Clinical, Research, and Direct-to-Consumer Contexts.
37. Agencia Española de Protección de Datos. Información sobre los países que tienen reconocido un nivel adecuado de protección o sobre los requisitos de las transferencias internacionales de datos. Disponible en: https://www.agpd.es/portalwebAGPD/canalresponsable/transferencias_internacionales/index-ides-idphp.php
38. Building on our inheritance. A report by the Human Genomics Strategy Group, 2012.
39. Modelo de uso e intercambio de datos. Deciphering Developmental Disorders (DDD). Disponible en: <http://www.ddduk.org/updates.html>
40. International Cancer Genome Consortium, Hudson TJ, Anderson W, Artez A, Barker AD, Bell C, Bernabé RR, et al. International network of cancer genome projects. *Nature*. 2010 Apr 15;464(7291):993-8. doi: 10.1038/nature08987. Erratum in: *Nature*. 2010 Jun 17;465(7300):966. Himmelbaue, Heinz [corrected to Himmelbauer, Heinz]; Gardiner, Brooke A [corrected to Gardiner, Brooke B]; Cross, Anthony [corrected to Cros, Anthony].
41. Zhang J, Baran J, Cros A, Guberman JM, Haider S, Hsu J, et al. International Cancer Genome Consortium Data Portal—a one-stop shop for cancer genomics data. *Database (Oxford)*. 2011;2011:bar026.
42. Joint Committee on Medical Genetics (2011) Consent and confidentiality in clinical genetic practice: Guidance on genetic testing and sharing genetic information. London: Royal College of Physicians.
43. Weitzel JN, Blazer KR, MacDonald DJ, Culver JO, Offit K. Genetics, genomics, and cancer risk assessment: State of the art and future directions in the era of personalized medicine. *CA Cancer J Clin*. 2011 Sep;61(5):327-59. doi: 10.1007/s00439-011-1028-3.
44. Offit K. Personalized medicine: new genomics, old lessons. *Hum Genet*. 2011 Jul;130(1):3-14. doi: 10.1007/s00439-011-1028-3.
45. Rodríguez-Santiago B, Armengol L. Tecnologías de secuenciación de nueva generación en diagnóstico genético pre- y postnatal. *Diagn Prenat*. 2012;23(2):56-66.
46. Wetter Strand KA. ADN Sequencing Costs: Data from the NHGRI Genome Sequencing Program (GSP) [Acceso 14 de enero de 2015]. Disponible en: www.genome.gov/sequencingcosts.
47. Sboner A, Mu XJ, Greenbaum D, Auerbach RK, Gerstein M. The real cost of sequencing: higher than you think! *Genome Biol*. 2011;12:125. doi: 10.1186/gb-2011-12-8-125.

LEGISLACIÓN

INVESTIGACIÓN BIOMÉDICA

NACIONAL

- Ley 14/2007, de 3 de julio de Investigación Biomédica.
- Real Decreto 1716/2011, de 18 de noviembre, por el que se establecen los requisitos básicos de autorización y funcionamiento de los biobancos con fines de investigación biomédica y del tratamiento de las muestras biológicas de origen humano, y se regula el funcionamiento y organización del Registro Nacional de Biobancos para investigación biomédica.

AUTONÓMICA

- Comunidad Autónoma de Andalucía: Decreto 1/2013, de 8 de enero, por el que se regula la autorización para la constitución y funcionamiento de biobancos con fines de investigación biomédica, se crean el Registro de Biobancos de Andalucía y el Biobanco del Sistema Sanitario Público de Andalucía.
- Comunidad Valenciana: Decreto 143/2008, de 3 de octubre, del Consell, por el que se regulan los biobancos en la Comunitat Valenciana.
- Comunidad Autónoma de Galicia: Ley 3/2005, de 7 de marzo de modificación de la Ley 3/2001, de 28 de mayo, reguladora del consentimiento informado y de la historia clínica de los pacientes.
- Comunidad Autónoma del País Vasco: Decreto 135/2015, de 7 de julio, sobre el régimen de autorización y funcionamiento de biobancos con fines de investigación biomédica en la Comunidad Autónoma de Euskadi.

PROTECCIÓN DE DATOS

- Ley Orgánica 15/1999, de 13 de diciembre, de protección de datos de carácter personal.
- Real Decreto 1720/2007, de 21 de diciembre, por el que se aprueba el Reglamento de

desarrollo de la Ley Orgánica 15/1999, de 13 de diciembre, de protección de datos de carácter personal.

DERECHOS DE LOS PACIENTES Y DOCUMENTACIÓN CLÍNICA

NACIONAL

- Ley 41/2002, de 14 noviembre, básica reguladora de la autonomía del paciente y de derechos y obligaciones en materia de información y documentación clínica (BOE 15.11.02).
- Orden SSI/2065/2014 de cartera de servicios comunes del Sistema Nacional de Salud y el procedimiento para su actualización, incluyendo las pruebas genéticas (31 de octubre de 2014).

AUTONÓMICA

- Aragón. Ley 6/2002, de 15 de abril, de Salud.
- Andalucía. Ley 2/1998 de 15 de junio de Salud.
- Asturias. Ley 1/1992, de 2 de julio, del Servicio de Salud del Principado de Asturias.
- Baleares. Ley 5/2003 de 4 de abril, de Salud de les Illes Balears.
- Canarias. Decreto 178/2005, de 26 de julio, por el que se aprueba el Reglamento que regula la historia clínica en los centros y establecimientos hospitalarios y establece el contenido, conservación y expurgo de sus documentos.
- Cantabria. Ley 7/2002, 10 de diciembre, de Ordenación Sanitaria de Cantabria.
- Castilla La Mancha. Ley 8/2000, de 30 de noviembre, de Ordenación Sanitaria de Castilla-La Mancha.
- Castilla y León. Ley 8/2003, de 8 de abril, sobre derechos y deberes de las personas en relación con la salud.
- Cataluña. Ley 21/2000, de 29 de diciembre, sobre los derechos de información concerniente a la salud y la autonomía del paciente.
- Extremadura. Ley 3/2005, de 8 de julio, de información sanitaria y autonomía del paciente.
- Galicia. Ley 8/2008, de 10 de julio, de salud de Galicia.
- La Rioja. Ley 2/2002, de 17 de abril, de Salud.
- Madrid. Ley 12/2001, de 21 de diciembre, de Ordenación Sanitaria de la Comunidad de Madrid.
- Murcia. Ley 4/1994, de 26 de julio de salud.
- Navarra. Ley Foral 11/2002, de 6 de mayo, sobre los derechos del paciente a las voluntades anticipadas, a la información y a la documentación clínica.
- País Vasco. Decreto 38/2012, de 13 de marzo, sobre historia clínica y derechos y obligaciones de pacientes y profesionales de la salud en materia de documentación clínica.
- Valencia. Ley 1/2003, de 28 de enero, de la Comunidad Valenciana.

BIOSEGURIDAD EN LA CIRCULACIÓN INTERNACIONAL DE MUESTRAS

- Real Decreto 65/2006, de 30 de enero, por el que se establecen requisitos para la importación y exportación de muestras biológicas.

MODELOS DE HOJAS DE INFORMACIÓN Y CONSENTIMIENTO

Modelo para diagnóstico (1 de 1)

Hoja de información al paciente para la realización de pruebas genéticas diagnósticas (1)

Solicitamos su consentimiento para:

1. Realizar pruebas genéticas de laboratorio en muestras biológicas (*sangre/tejidos/otros fluidos biológicos*) cuya finalidad es confirmar su diagnóstico clínico/realizar un estudio predictivo/de portador/diagnosticar si está afectado o es portador de (p. ej., una *enfermedad metabólica hereditaria*).

En el informe se detallarán los procedimientos técnicos que se lleven a cabo y si se utilizan técnicas de secuenciación completa del genoma, así como la capacidad diagnóstica de la prueba.

2. Dichas pruebas serán realizadas en el Servicio de del Hospital o en otros laboratorios externos nacionales o extranjeros en el caso de que la metodología diagnóstica requerida no se realice actualmente en este Servicio.
3. Únicamente el personal sanitario debidamente autorizado podrá acceder a los datos personales y a los resultados de las pruebas genéticas.
4. El facultativo que le solicita estas pruebas adquiere el compromiso de suministrar información acerca del objeto de los análisis y facilitarle asesoramiento genético, lo que incluirá la comunicación de las diferentes opciones de actuación cuando se conozcan los resultados.
5. En las pruebas, podría ser que se encontrase información no directamente relacionada con el objeto de los análisis y usted puede decidir si desea o no que se le comunique.
6. La información obtenida puede ser relevante también para sus familiares, y en este caso le explicaremos el porqué de la conveniencia de que la conozcan. Es decisión personal suya informarles, algo que nosotros le recomendamos, con el fin de que, si ellos lo desean, puedan acudir a una consulta especializada en genética donde les informarán sobre su riesgo personal y sus opciones de salud en el futuro.

En caso de que usted haya optado por no conocer la información resultante de los análisis, se valorará la necesidad de comunicársela por nuestra parte a sus familiares. Le solicitamos que nos facilite la identidad de una persona de contacto a estos efectos.
7. Una vez finalizados los análisis, los datos obtenidos y las muestras excedentes se guardarán en el Servicio del Hospital, por el interés que pueden tener para satisfacer futuras necesidades asistenciales de usted y sus familiares.

En caso de que se pretenda utilizar sus muestras o sus datos con fines de investigación biomédica, le informaremos sobre ello y le solicitaremos su consentimiento. (*Si el destino de las muestras sobrantes es su destrucción, debe modificarse convenientemente el redactado de este punto*).

Firma del paciente

Firma del facultativo

(1) Este modelo se ha basado en el propuesto por la Comisión de Ética de la AEGH y aprobado por su Junta Directiva. Disponible en: <http://www.elsevier.es/es-revista-diagnostico-prenatal-327-articulo-el-documento-consentimiento-informado-realizacion-90200639>. También se han utilizado como base los modelos del Hospital 12 de Octubre.

Para enfocar el modelo hacia estudios de secuenciación masiva se ha tenido en cuenta: Ayuso C, Millán JM, Mancheño M, Dal-Ré R. Informed consent for whole-genome sequencing studies in the clinical setting. Proposed recommendations on essential content and process. Eur J Hum Genet. 2013;21:1054-9.

Modelo Donación de muestras biológicas a biobanco (1 de 3)

Donación de muestras biológicas al biobanco

1. ¿Para qué dono mi muestra?

Para avanzar en el conocimiento de las enfermedades y en su prevención, diagnóstico y tratamiento es necesario disponer de datos clínicos y muestras biológicas de pacientes y donantes sanos para analizarlos y obtener conclusiones (1). La donación de la muestra implica su utilización para contribuir en este sentido al avance de la ciencia. No debe esperar un beneficio directo para su salud a partir de esta donación, **sino participar en el avance general del conocimiento científico.**

2. ¿Qué muestras me solicitan? ¿La donación representa riesgos físicos?

Se solicita la donación de sangre y tejido sobrante de las pruebas que, como parte del actual proceso asistencial, se le han realizado o se le van a realizar en este centro (1)/Se le solicita la extracción y donación de xx de sangre. La extracción no le causará más molestias que las ocasionadas por una extracción de sangre habitual, y pueden incluir la aparición de pequeños hematomas en la zona de punción que desaparecen transcurridos uno o dos días (2).

3. ¿Dónde y en qué condiciones se van a almacenar las muestras y los datos asociados?

Las muestras quedarán almacenadas en el **biobanco** que ha sido autorizado por el Departamento de Sanidad..... y está inscrito en el Registro Nacional de Biobancos. Los biobancos son instituciones sin ánimo de lucro con vocación de **servicio público** cuyo objetivo es facilitar la disponibilidad de muestras para la comunidad científica.

Las muestras se almacenarán asociadas a un **código** que solo podrá ser relacionado con su identidad por el personal autorizado del biobanco, nunca por los investigadores que las utilicen en sus proyectos.

Los datos que directa o indirectamente estén relacionados con su identidad se almacenarán en el biobanco cumpliendo las **medidas de seguridad** previstas en la legislación sobre protección de datos de carácter personal. El biobanco es el responsable de este fichero de datos, ante el que podrá ejercer los derechos de acceso, rectificación y cancelación y oposición.

4. ¿Se accederá a otros datos relacionados con mi salud?

Le solicitamos su **consentimiento para poder consultar su historia clínica** cuando un investigador lo solicite por ser necesario para la investigación, y siempre que el Comité de Ética del Biobanco lo apruebe (1). El biobanco ha acordado con el centro hospitalario un mecanismo controlado y seguro para llevar a cabo este acceso.

5. ¿Quién, para qué y en qué condiciones se utilizarán las muestras y los datos?

Las muestras y los datos se cederán a investigadores **para proyectos aprobados** en sus centros, en el marco de las finalidades para las que ha consentido, después de un examen de la solicitud por los comités científico y de ética externos del biobanco. Si el comité de ética del biobanco lo estimara adecuado por las características específicas de un proyecto, nos pondríamos en contacto con usted para solicitar su consentimiento para esa cesión en particular.

Los investigadores recibirán las muestras y los datos codificados (si la investigación requiere mantener la trazabilidad con sus datos clínicos) o anonimizadas (destruyendo el código de identificación). Nunca recibirán información que le pueda identificar y adquieren un **compromiso legal con el biobanco para la correcta utilización de la muestra y los datos.**

Estas cesiones pueden tener como destino proyectos de **cualquier país del mundo**, incluyendo países sin las exigencias para la protección de datos personales que las previstas en la legislación española. Sin embargo, el biobanco exigirá a los investigadores extranjeros y a sus instituciones **los mismos compromisos que a los investigadores españoles.** Las transferencias internacionales de sus datos se realizarán con las garantías que exige la legislación española.

Los investigadores utilizarán diferentes técnicas para analizar sus muestras, incluida la **secuenciación de todo su genoma.**

Modelo Donación de muestras biológicas a biobanco (2 de 3)

Los datos podrán ser **publicados** en revistas científicas en condiciones que aseguren que el donante no será identificado. También podrán ser depositados en **bases de datos** disponibles para toda la comunidad científica a través de procedimientos revisados por el comité de ética externo del biobanco (3).

La información sobre los proyectos en que se han utilizado muestras del biobanco está disponible en su página web. Además, si usted lo solicita dirigiéndose al biobanco por correo postal o electrónico o fax, se le facilitará información sobre aquellos proyectos en particular en los que se hayan utilizado sus muestras o sus datos.

6. ¿Se va a obtener información sobre mi salud o la de mis familiares? ¿Se me va a comunicar? ¿Qué relevancia tendrá?

Los métodos utilizados en investigación biomédica suelen ser diferentes de los aprobados para la práctica clínica, por lo que no deben de ser considerados con valor clínico para usted. Sin embargo, en el caso de que estas investigaciones proporcionen datos que pudieran ser clínica o genéticamente relevantes para usted e interesar a su salud o a la de su familia, **le serán comunicados salvo que indique lo contrario** en la casilla que aparece al final de este documento. También le corresponderá a usted **decidir si quiere o no comunicar esta información a sus familiares**.

Si usted no desea recibir esta información, tenga en cuenta que la ley establece que, cuando la información obtenida sea necesaria para evitar un grave perjuicio para la salud de sus familiares biológicos, un comité de expertos estudiará el caso y deberá decidir si es conveniente informar a los afectados o a sus representantes legales (1).

La comunicación de esta información se llevará a cabo por **profesionales que le podrán explicar** adecuadamente su relevancia y las opciones que se pudieran plantear.

7. ¿Puedo acceder yo a los resultados de los análisis si lo solicito?

Si usted lo solicita **podrá acceder** a cualquier información que se obtenga a partir de los análisis de sus muestras contactando con el biobanco. Los datos de contacto figuran al final de este documento.

8. ¿Hasta cuándo se van a guardar la muestra y los datos? ¿Y si decido que ya no quiero que mi muestra o datos se utilicen más? ¿Y si necesito la muestra?

Las muestras y los datos se almacenarán en el biobanco y en las bases de datos de resultados de investigación mientras **puedan ser útiles para la investigación biomédica**. Si el biobanco cesara en su actividad, la información sobre el destino de las muestras y los datos estará disponible en el **Registro Nacional de Biobancos**.

Podrá solicitar en cualquier momento y sin necesidad de justificación revocar su consentimiento para la donación, dirigiéndose al biobanco. Usted decidirá entonces si las muestras se destruyen, o se mantienen pero sin posibilidad de relacionarlas con su identidad (anonimización). Los datos que se hayan obtenido hasta entonces en los proyectos de investigación a partir del análisis de sus muestras no se eliminarán, pero ya no podrán ser asociados con su identidad.

La muestra estará disponible para usted si la necesitara por motivos de salud mientras no se agote en el curso de las investigaciones.

9. ¿La donación me reportará algún beneficio económico?

Usted no obtendrá ni ahora ni en el futuro **ningún beneficio económico** por esta donación, ni tendrá derechos sobre posibles beneficios comerciales de los descubrimientos que puedan conseguirse como resultado de la investigación biomédica. Sin embargo, los conocimientos obtenidos gracias a los estudios llevados a cabo a partir de su muestra y de muchas otras pueden ayudar al avance médico y, por ello, a otras personas (1).

10. ¿Qué regulación protege los derechos del donante?

Son especialmente relevantes: la Ley 14/2007 de Investigación Biomédica y la Ley Orgánica 15/1999 de Protección de Datos. Las condiciones de la donación, almacenamiento, cesión y utilización de las muestras y los datos se llevará a cabo cumpliendo lo previsto en estas disposiciones.

Modelo Donación de muestras biológicas a biobanco (3 de 3)

En resumen: El biobanco es una institución autorizada por la Administración Pública que actúa como intermediaria entre el donante y los investigadores. Almacena las muestras y los datos en condiciones seguras y los distribuye para proyectos científicamente relevantes llevados a cabo por investigadores solventes que se comprometan a su uso legítimo. Pone a disposición de los donantes toda la información sobre su utilización.

Datos del Biobanco:

Titular:

Director científico:

Página web:

Dirección:

Teléfono:

Correo electrónico de contacto:

Número de Registro:

**Casillas para señalar opciones en relación con la comunicación de resultados
y con la restricción de utilización de las muestras.**

Fecha y firmas del donante y del representante del biobanco

1. Del modelo de la Red Nacional de Biobancos.

http://www.redbiobancos.es/DownloadHandler.ashx?f=Modelo_de_Consentimiento_Informado_de_la_Red_Nacional.pdf&s=34&p=128&d=319

2. Del modelo de consentimiento informado para la utilización de datos clínicos y material biológico para investigación biomédica y su conservación en el Biobanc del Hospital Clínic-IDIBAPS.

3. Se debería valorar si incluir alguna advertencia sobre la posibilidad remota de identificación. La hoja de información del proyecto 1000 genomas señala::

This information will be put in open access scientific databases, available on the Internet to anyone who wants to look at it. Although only experts will know how to interpret this information, there is a small chance that somebody could figure out how to connect you with the information from the study of the sample you give; the information could then be used to discriminate against you or your family members. Currently, we believe this could happen only if somebody knew that you had given a sample to be studied for this project and:

- *got another sample from you, found an expert to test that sample, and then somebody knew that you had given a sample to be studied for this project and;*
- *got another sample from you, found an expert to test that sample, and then compared the genetic information from that test with the genetic information in the scientific databases;*
- *found an expert to compare the genetic information about you in the scientific databases with information known to have come from you (or from a family member) included in some other database developed by someone else for some other purpose; or*
- *found an expert to look in the scientific databases for a particular genetic variation known (or someday found) to be associated with a disease or trait that you have or carry, that others know about or can see, and that is very rare.*

Any of these things would require that the person trying to link the information to you knew that you participated in the project. For this reason, to minimize these risks, you may wish to limit the number of people you tell about your participation.

As technology advances, there may be new ways of linking information back to you that we cannot foresee now. Also, we cannot always foresee the results of research, so new risks may come up in the future that we cannot predict now. We believe that the benefits of learning more about human genetic variation and how it relates to health and disease outweigh the current and potential future risks, but this is something that you must judge for yourself. If you believe you have been injured because of this research, please contact [contact person]. Decisions about payment for medical treatment for injuries relating to your participation in research will be made by [institution].

Modelo Donación de muestras biológicas a proyecto y colección o biobanco (1 de 4)

**Donación de muestras biológicas para el proyecto
y la colección /biobanco**

1. ¿Para qué dono mi muestra?

Se solicitan sus muestras para ser utilizadas en el **PROYECTO**
(incluir la referencia a su financiación), cuyo objetivo es

Los beneficios que se derivarán del proyecto son

El proyecto se va a desarrollar en

El investigador principal es

También participan

El proyecto ha sido aprobado por el Comité de Ética de que ha revisado que es conforme a la legislación española y a los principios éticos que deben ser respetados en la investigación biomédica.

Cuando termine el proyecto las muestras quedarán almacenadas en la **COLECCIÓN**
..... con la finalidad de ser utilizadas para investigaciones relacionadas con
..... y cuyo responsable es La colección se registrará en el Registro Nacional de Biobancos (Sección colecciones)/en el **BIOBANCO**

2. ¿Qué muestras me solicitan? ¿La donación representa riesgos físicos?

Se solicita la donación de sangre y tejido sobrante de las pruebas que, como parte del actual proceso asistencial, se le han realizado o se le van a realizar en este centro / Se le solicita la extracción y donación de xx de sangre. La extracción no le causará más molestias que las ocasionadas por una extracción de sangre habitual, y pueden incluir la aparición de pequeños hematomas en la zona de punción que desaparecen transcurridos uno o dos días.

3. ¿Dónde y en qué condiciones se van a analizar y almacenar las muestras y los datos asociados durante el proyecto?

Durante el desarrollo del proyecto sus datos y sus muestras se analizarán en los centros participantes. Usted puede conocer los centros en la web Las muestras se almacenarán asociadas a un **código** que solo podrá ser relacionado con su identidad por el personal autorizado. Los datos que directa o indirectamente estén relacionados con su identidad se utilizarán cumpliendo las **medidas de seguridad** previstas en la legislación sobre protección de datos de carácter personal. El responsable de sus datos es

4. A. ¿Quién, para qué y en qué condiciones utilizarán las muestras y los datos almacenados en la colección?

Las muestras y los datos que se almacenen en la colección se utilizarán en **proyectos en que participe el investigador responsable**, relacionados con
..... y **aprobados** por un comité de ética. Si el comité lo estima adecuado por las características específicas de un proyecto, nos pondremos en contacto con usted para solicitar su consentimiento para ese proyecto en particular.

Las muestras se almacenarán asociadas a un **código** que solo podrá ser relacionado con su identidad por el personal autorizado. Los datos que directa o indirectamente estén relacionados con su identidad se utilizarán cumpliendo las **medidas de seguridad y las garantías jurídicas que impidan su identificación** previstas en la legislación sobre protección de datos de carácter personal. El responsable de sus datos será

Modelo Donación de muestras biológicas a proyecto y colección o biobanco (2 de 4)

En estos proyectos podrán participar investigadores de centros españoles o de **cualquier país del mundo**, incluyendo países con menos exigencias para la protección de datos personales que las previstas en la legislación española. Sin embargo, se exigirá a todos los investigadores y a sus instituciones **los mismos compromisos previstos en la legislación española en relación con la seguridad en el almacenamiento de los datos y las muestras y con su utilización legítima**.

En estos proyectos se utilizarán diferentes técnicas para analizar sus muestras, incluida la **secuenciación de todo su genoma**.

Los datos podrán ser publicados en revistas científicas y depositados en **bases de datos disponibles para toda la comunidad científica**, siempre en condiciones que aseguren su uso legítimo y que el donante no será identificado. Estas condiciones serán **controladas y revisadas por el Comité de Ética de** (3).

La información sobre los proyectos en que se utilicen sus muestras y sus datos le será facilitada si usted lo solicita dirigiéndose al responsable de la colección cuyos datos figuran al final de este documento.

4. **B. ¿Quién, para qué y en qué condiciones utilizarán las muestras y los datos almacenados en el biobanco?** (En este caso ya no procederá añadir lo que se indica a continuación en las preguntas 5 a 10, redactado para el supuesto en que el destino sea una colección).

5. **¿Se accederá a otros datos relacionados con mi salud?**

Le solicitamos su **consentimiento para poder consultar su historia clínica** con la finalidad de conocer datos necesarios para los objetivos de este proyecto y de los que se lleven a cabo en el futuro en el marco de la colección Únicamente el personal autorizado en el hospital tendrá acceso a su historia. Los datos podrán ser cedidos a investigadores participantes en el proyecto de forma que no puedan conocer su identidad y para los objetivos de esta investigación.

6. **¿Se va a obtener información sobre mi salud o la de mis familiares? ¿Se me va a comunicar? ¿Qué relevancia tendrá?**

Los métodos utilizados en investigación biomédica suelen ser diferentes de los aprobados para la práctica clínica, por lo que no deben de ser considerados con valor clínico para usted. Sin embargo, en el caso que este proyecto o los que se desarrollen en el futuro proporcionen datos que pudieran ser clínica o genéticamente relevantes para usted e interesar a su salud o a la de su familia, **le serán comunicados si así lo indica** en la casilla que aparece al final de este documento. También le corresponderá a usted **decidir si quiere o no comunicar esta información a sus familiares**.

Si usted no desea recibir esta información, tenga en cuenta que la ley establece que, cuando la información obtenida sea necesaria para evitar un grave perjuicio para la salud de sus familiares biológicos, un comité de expertos estudiará el caso y deberá decidir si es conveniente informar a los afectados o a sus representantes legales (1).

La comunicación de esta información se llevará a cabo por **profesionales que le podrán explicar** adecuadamente su relevancia y las opciones que se pudieran plantear.

7. **¿Puedo acceder yo a los resultados de los análisis si lo solicito?**

Si usted lo solicita **podrá acceder** a cualquier información que se obtenga a partir de los análisis de sus muestras contactando con el responsable de la colección.

8. **¿Hasta cuándo se van a guardar la muestra y los datos? ¿Y si decido que ya no quiero que mi muestra o datos se utilicen más? ¿Y si necesito la muestra?**

Las muestras y los datos se almacenarán en la colección y en las bases de datos de resultados de investigación mientras **puedan ser útiles para la investigación** en Si el investigador responsable cesara en su actividad, la información sobre el destino de las muestras y los datos estará disponible en

Modelo Donación de muestras biológicas a proyecto y colección o biobanco (3 de 4)

Podrá solicitar en cualquier momento y sin necesidad de justificación revocar su consentimiento para la donación, dirigiéndose al responsable del proyecto y cuando este haya finalizado, al responsable de la colección. Usted decidirá entonces si las muestras se destruyen, o se mantienen pero sin posibilidad de relacionarlas con su identidad (anonimización). Los datos que se hayan obtenido hasta entonces en los proyectos de investigación a partir del análisis de sus muestras no se eliminarán, pero ya no podrán ser asociados a su identidad en ningún caso.

La muestra estará disponible para usted si la necesitara por motivos de salud mientras no se agote en el curso de las investigaciones.

9. ¿La donación me reportará algún beneficio económico?

Usted no obtendrá ni ahora ni en el futuro **ningún beneficio económico** por esta donación, ni tendrá derechos sobre posibles beneficios comerciales de los descubrimientos que puedan conseguirse como resultado de la investigación biomédica. Sin embargo, los conocimientos obtenidos gracias a los estudios llevados a cabo a partir de su muestra y de muchas otras pueden ayudar al avance médico y, por ello, a otras personas (1).

10. ¿Qué regulación protege los derechos del donante?

Son especialmente relevantes: la Ley 14/2007 de Investigación Biomédica y la Ley Orgánica 15/1999 de Protección de Datos. Las condiciones de la donación, almacenamiento, cesión y utilización de las muestras y los datos se llevará a cabo cumpliendo lo previsto en estas disposiciones.

En resumen: usted dona las muestras y los datos para participar en el proyecto Cuando el proyecto termine las muestras pasarán a formar parte de la colección/Biobanco, donde se almacenarán en condiciones seguras y se utilizarán para proyectos científicamente relevantes llevados a cabo por Toda la información sobre su utilización estará a su disposición.

Casillas para señalar opciones en relación con la comunicación de resultados y con la restricción de utilización de las muestras.

Modelo Donación de muestras biológicas a proyecto y colección o biobanco (4 de 4)

Datos de la colección/Biobanco:

Investigador responsable/Titular y director científico:

Centro/Página web:

Dirección:

Teléfono:

Correo electrónico de contacto:

Número de Registro:

1. Del modelo de la Red Nacional de Biobancos. http://www.redbiobancos.es/DownloadHandler.ashx?f=Modelo_de_Consentimiento_Informado_de_la_Red_Nacional.pdf&s=34&p=128&d=319

2. Del modelo de consentimiento informado para la utilización de datos clínicos y material biológico para investigación biomédica y su conservación en el Biobanc del Hospital Clínic-IDIBAPS.

3. Se debería valorar si incluir alguna advertencia sobre la posibilidad remota de identificación. La hoja de información del proyecto 1000 genomas señala:

This information will be put in open access scientific databases, available on the Internet to anyone who wants to look at it. Although only experts will know how to interpret this information, there is a small chance that somebody could figure out how to connect you with the information from the study of the sample you give; the information could then be used to discriminate against you or your family members. Currently, we believe this could happen only if somebody knew that you had given a sample to be studied for this project and:

- *got another sample from you, found an expert to test that sample, and then somebody knew that you had given a sample to be studied for this project and;*
- *got another sample from you, found an expert to test that sample, and then compared the genetic information from that test with the genetic information in the scientific databases;*
- *found an expert to compare the genetic information about you in the scientific databases with information known to have come from you (or from a family member) included in some other database developed by someone else for some other purpose; or*
- *found an expert to look in the scientific databases for a particular genetic variation known (or someday found) to be associated with a disease or trait that you have or carry, that others know about or can see, and that is very rare.*

Any of these things would require that the person trying to link the information to you knew that you participated in the project. For this reason, to minimize these risks, you may wish to limit the number of people you tell about your participation.

As technology advances, there may be new ways of linking information back to you that we cannot foresee now. Also, we cannot always foresee the results of research, so new risks may come up in the future that we cannot predict now. We believe that the benefits of learning more about human genetic variation and how it relates to health and disease outweigh the current and potential future risks, but this is something that you must judge for yourself. If you believe you have been injured because of this research, please contact [contact person]. Decisions about payment for medical treatment for injuries relating to your participation in research will be made by [institution].