

COLABORACIÓN ESPECIAL

Recibido: 26 de octubre de 2021
Aceptado: 23 de noviembre de 2021
Publicado: 4 de febrero de 2022

CRIBADO NEONATAL GENÓMICO. PERSPECTIVA DE LA COMISIÓN DE ÉTICA DE LA ASOCIACIÓN ESPAÑOLA DE GENÉTICA HUMANA. PARTE I. LAS TECNOLOGÍAS DE SECUENCIACIÓN MASIVA (NGS) Y SU APLICACIÓN AL CRIBADO NEONATAL. DESAFÍOS Y OPORTUNIDADES

Teresa Pàmpols Ros (1,6), Antonio Pérez Aytés (2,6), José Miguel García Sagredo (3,6), Aránzazu Díaz de Bustamante (4,6) e Ignacio Blanco Guillermo (5,6)

- (1) Sección de errores congénitos del metabolismo-IBC. Servicio de Bioquímica y Genética Molecular. Hospital Clínico de Barcelona. Barcelona. España.
(2) Grupo de Investigación en Perinatología. Instituto de Investigación Sanitaria. Hospital La Fe. Valencia. España.
(3) Facultad de Medicina. Universidad de Alcalá. Alcalá de Henares (Madrid). España.
(4) Unidad de Genética. Hospital Universitario de Móstoles. Móstoles (Madrid). España.
(5) Hospital Universitario Germans Trias i Pujol. Badalona (Barcelona). España.
(6) Comisión de Ética de la Asociación Española de Genética Humana (AEGH). España.

Los autores declaran que no existe ningún conflicto de intereses.

RESUMEN

En 2003, cuando finalizó el Proyecto Genoma Humano, surgió la idea de secuenciar el genoma a todos los recién nacidos, archivarlo en la historia clínica y usarlo a lo largo de toda la vida para manejar riesgos de enfermedades y respuesta a medicamentos. Dieciocho años más tarde, las promesas de la medicina genómica y el extraordinario abaratamiento de las tecnologías secuenciadoras, siguen alimentando este sueño que todavía plantea grandes desafíos prácticos, éticos y sociales y la secuenciación genómica se presenta como el próximo gran cambio histórico en los programas de cribado neonatal.

En el presente artículo, se analizan los retos y oportunidades de las tecnologías secuenciadoras de nueva generación, sus costos reales, la problemática inherente a la gestión, almacenamiento y protección de la enorme cantidad de datos genómicos que generan y finalmente, en base a las conclusiones de investigaciones recientes, se considera el potencial y limitaciones de su aplicación en dos escenarios, el recién nacido enfermo con finalidades diagnósticas y el recién nacido sano, asintomático, con finalidades de salud pública (programas de cribado neonatal). En una segunda parte de este artículo se tendrán en cuenta los aspectos éticos, legales y sociales (AELS).

El objetivo final es contribuir al debate científico, profesional, ético y social, promoviendo que la secuenciación genómica en el recién nacido no sea usada indiscriminadamente constituyendo un riesgo, sino que bien empleada sea una aliada en la promoción de la salud y prevención de las consecuencias de las enfermedades genéticas.

Palabras clave: Cribado neonatal, Cribado neonatal genómico, Secuenciación del recién nacido, Cribado genómico, Cribado genético, Tecnologías de secuenciación masiva, Secuenciación del genoma, Secuenciación del exoma, Salud pública, Salud pública genómica.

Correspondencia:
Teresa Pàmpols
Sección de errores congénitos del metabolismo-IBC
Servicio de bioquímica y genética molecular
Hospital Clínico de Barcelona, sede Maternidad
C/ Mejía Lequerica, s/n, edificio Helios III, planta baja
08028 Barcelona, España
tpampols@clinic.cat

ABSTRACT

Genomic newborn screening. Perspective from the Ethics commission of the Spanish Society for Human Genetics. Part I. Next generation sequencing technologies applied to newborn screening. Challenges and opportunities

In 2003 at the ending of the Human Genome Project, it aroused the idea that all newborns could be sequenced and its genome archived in the clinical record, in order to manage risks of diseases and response to medications along his whole life. Eighteen years later, promises of genomic medicine and tremendous decrease of costs of next generation sequencing technologies, continues feeding this dream that shows important practical, ethical and social challenges and genomic sequencing is presented as the next historical change in newborn screening programs.

In this article we analyze challenges and opportunities of next generation sequencing technologies, their real costs, problems associated to management, storage and protection of the enormous amount of genomic data produced and finally, according to conclusions of recent researches, there are considered the conclusions in two contexts, sick newborn with diagnostic purposes and healthy asymptomatic newborns with public health purposes (newborn screening programs). In a second part of this article it will be considered ethical, legal and social issues (ELSI).

Final objective is to contribute to scientific, professional, ethics and social debate in order to promote that genome sequencing in newborn don't be used indiscriminately constituting a risk, but properly done, as a partner in the promotion of health and prevention of consequences of genetic diseases.

Key words: Newborn screening, Genomic newborn screening, Newborn sequencing, Genomic screening, Genetic screening, Next generation sequencing technologies, Genome sequencing, Exome sequencing, Public health, Genomic public health.

Cita sugerida: Pàmpols Ros T, Pérez Aytés A, García Sagredo JM, Díaz de Bustamante A, Blanco Guillermo I. Cribado neonatal genómico. Perspectiva de la Comisión de ética de la Asociación Española de Genética Humana. Parte I. Las tecnologías de secuenciación masiva (NGS) y su aplicación al cribado neonatal. Desafíos y oportunidades. Rev Esp Salud Pública. 2022; 96: 4 de febrero e202202012.

INTRODUCCIÓN

El proyecto genoma humano fue un esfuerzo colaborativo internacional de 13 años (1990-2003), cuyo principal objetivo fue determinar la secuencia completa de bases del ADN e identificar y cartografiar los genes de un genoma humano promedio.

Desde entonces los conocimientos genómicos han invadido todos los campos de la medicina y de la investigación biomédica y han hecho emerger una nueva disciplina médica, la medicina genómica, que contempla el uso de la información genómica de un individuo como parte de sus cuidados clínicos (diagnóstico, decisiones terapéuticas). La medicina genómica es una parte fundamental de la medicina de precisión en la que los datos genómicos, epigenómicos, de exposición medioambiental y otros datos de salud, mediante su tratamiento como big data y las herramientas de la inteligencia artificial, se usarán para guiar con mayor precisión el diagnóstico y tratamiento individual.

En 2003, el *National Health Service* (NHS) del Reino Unido lanzó la idea de que sería posible “secuenciar a los recién nacidos” y obtener un mapa completo de marcadores genéticos clave o incluso un genoma completo, de manera que podría archivar en la historia clínica y usarse a lo largo de toda la vida para manejar riesgos de enfermedades y respuesta a los medicamentos⁽¹⁾. De hecho, para los seres humanos, evitar las consecuencias del azar y ganar previsibilidad es un objetivo en todos sus ámbitos de actividad, economía, agricultura, clima, etc. y por supuesto en el ámbito de la genética y del control de las enfermedades.

Francis Collins que dirigió el Proyecto Genoma Humano y actualmente ejerce de director de los *National Institutes for Health* (NIH) de EE.UU., ha expresado en múltiples ocasiones la esperanza de que todos los recién

nacidos pudieran ser secuenciados al nacimiento, estableciendo con ello una etapa de cuidados médicos para toda la vida y acciones preventivas a medida del genoma de cada niño. Sería un punto álgido en el desarrollo de la medicina genómica de las 4 P: personalizada, predictiva, preventiva y participativa. Con ello se integraría de forma continuada el conocimiento genético a lo largo de nuestras vidas y produciría un aumento de conocimientos y el desarrollo de tratamientos para enfermedades raras poco conocidas y no tratables⁽²⁾.

Este sueño de los impulsores de la medicina genómica, técnicamente está cada vez más cerca gracias al espectacular descenso de los costes de secuenciar un genoma que pasó de los 9.263.072 dólares que costó obtener el primer borrador en 2000 a 702 dólares en 2020 (ver el portal del NHGRI (*US National Human Genome Research Institute*, en <http://www.genome.gov>), el punto de inflexión en el descenso de los costos se produjo en 2007 cuando se desarrollaron las tecnologías de secuenciación de segunda generación o secuenciación paralela masiva de elevado rendimiento también llamadas *next generation sequencing technologies* (NGS) y *high-throughput sequencing* (HTS).

La combinación del conocimiento del mapa del genoma humano y este desarrollo de las NGS es la que se espera tenga un impacto transformador en la medicina, que hasta ahora no se ha materializado como se esperaba quizás por haberla enfocado hacia las enfermedades comunes como la diabetes o las enfermedades coronarias cuya arquitectura genética ha resultado ser mucho más compleja de lo que se creía, pero las predicciones sí que se han hecho realidad en el ámbito de la comprensión y prevención de los trastornos mendelianos (4.377 en la actualidad, de acuerdo con la base de datos OMIM, el catálogo en línea de genes humanos y trastornos mendelianos: go.nature.com/omimdb) y ha abierto la puerta al manejo

verdaderamente personalizado de los pacientes con enfermedades raras que en un 80% son de base genética⁽³⁾.

Empujan a la secuenciación genómica de los recién nacidos, no solo las promesas de la medicina genómica sino también el demostrado potencial de la secuenciación genómica para el diagnóstico y cuidado de los niños con trastornos genéticos y la oportunidad para detectarlos incluso antes de su debut clínico⁽⁴⁾.

Pero el ámbito del diagnóstico no se corresponde con el de los programas de cribado neonatal (PCN), que son actuaciones de salud pública aplicados a población asintomática de bajo riesgo que no ha pedido atención médica a causa de síntomas clínicos. Dichos programas están dirigidos a la identificación presintomática de trastornos congénitos tratables cuidadosamente seleccionados, a fin de que una intervención médica inmediata en los niños que son encontrados positivos, reduzca la morbi-mortalidad y las posibles discapacidades asociadas.

Es, además, una población vulnerable que requiere protección, por lo que los PCN se rigen por los principios de la ética biomédica y de la salud pública, para actuar en el mejor interés del niño, buscando un balance positivo de beneficios y riesgos, siguiendo un modelo basado en la evidencia y contemplando los principios, relevantes y ampliamente reconocidos para los cribados genéticos, de proporcionalidad, respeto hacia las personas y equidad.

La Comisión de ética de la AEGH es consciente de que en estos momentos el cribado neonatal genómico se presenta como un cambio histórico para los programas actuales, ya que tiene el potencial para incrementar el número de enfermedades mendelianas a incluir en un orden de magnitud muy superior al que representó en su día la introducción de la espectrometría de masas en tándem (MS/MS).

Alimentan el optimismo científico la eficiencia de las nuevas tecnologías secuenciadoras y su espectacular abaratamiento de costos junto con las promesas de la medicina de precisión, pero la secuenciación genómica es solo una tecnología, y al igual que otras, aunque presenten evidentes beneficios no están exentas de efectos no deseados. El empleo de las NGS en el recién nacido plantea desafíos prácticos, financieros y éticos que lo hacen difícil e incluso improbable para algunos. A pesar de ello, a menudo se ha sugerido que la secuenciación del genoma en el recién nacido es inevitable, pero ni lo es, ni debemos asumir que ésta u otras nuevas tecnologías deban ser usadas indiscriminadamente en toda su extensión y en todos los casos. El respeto por el potencial y las promesas de la medicina de precisión nos obliga a una reflexión en profundidad para determinar cuáles serían las mejores aplicaciones de las NGS en los recién nacidos, teniendo en cuenta el estado actual de conocimientos y la indispensable consideración de los aspectos éticos, legales y sociales⁽⁵⁾.

Con el espíritu de contribuir a esta reflexión, proporcionar elementos para el necesario debate científico, profesional, ético y social y promover que la secuenciación genómica no constituya un riesgo sino una aliada en la mejora de estos programas de salud pública, el artículo se ha estructurado en dos partes, en la **Parte I. Las tecnologías de secuenciación masiva y su aplicación al cribado neonatal. Desafíos y oportunidades**, se contemplarán: las tecnologías de secuenciación masiva (NGS) y la interpretación de resultados o cómo convertir los datos de la secuencia del ADN en una información clínicamente útil, su potencial y limitaciones; los costos reales de la secuenciación; el desafío de la gestión, almacenamiento y protección del diluvio de datos que generan las NGS y finalmente el uso de las NGS en el recién nacido enfermo con finalidades diagnósticas y en el recién nacido sano con finalidades de salud pública (cribado neonatal) sus desafíos y oportunidades.

En la Parte II: Aspectos éticos, legales y sociales a tener en cuenta respecto a la introducción de las tecnologías de secuenciación masiva (NGS) en un programa de cribado neonatal de salud pública, se contemplarán los principios éticos de la ética médica (autonomía/respeto por la persona, beneficencia, no maleficencia, equidad y justicia) y los específicos de la salud pública (maximización de la salud de la población, eficiencia, proporcionalidad y transparencia) que rigen los programas de cribado neonatal; las implicaciones jurídicas incluyendo la definición en nuestra legislación de cribado genético, las pruebas genéticas en menores, la protección de datos genéticos/genómicos y el acceso a la secuencia en bruto del genoma y finalmente los aspectos sociales, incluyendo la influencia del imperativo tecnológico, los actores comerciales (pruebas de cribado neonatal de venta directa a los consumidores), los grupos de apoyo de pacientes y finalmente la perspectiva de los padres y aspectos psicosociales.

LAS TECNOLOGÍAS DE SECUENCIACIÓN MASIVA (NGS) Y LA INTERPRETACIÓN DE RESULTADOS O COMO CONVERTIR LOS DATOS DE LA SECUENCIA DEL ADN EN UNA INFORMACIÓN CLÍNICAMENTE ÚTIL. POTENCIAL Y LIMITACIONES

A continuación, se describe brevemente la técnica de secuenciación genómica, a fin de dar a conocer su potencial y limitaciones, así como examinar el grado de cumplimiento de los tres primeros criterios para la evaluación de cualquier prueba genética que son su validez analítica (sensibilidad, especificidad), su validez clínica y su utilidad clínica^(6,7). El cuarto criterio, que es el respeto por los aspectos éticos, legales y sociales asociados (AELS), se comentará en la Parte II del artículo.

La doble hélice del ADN está formada por dos hebras constituidas por glucosa fosfato, ácido desoxiribonucleico y 4 tipos de bases nucleotídicas, adenina (A), citosina (C), timina (T) y guanina (G). La secuenciación del genoma es el proceso de determinar el orden de los millones de pares de bases en el ADN de un organismo. El ADN de una célula reproductora contiene 3.000 millones de pares de bases nucleotídicas, cada una de las restantes células del cuerpo humano 6.000 millones de pares de bases porque contiene el genoma que hemos recibido del padre y de la madre.

Las tecnologías actuales (NGS o *next generation sequencing technologies*) que permiten la secuenciación rápida del ADN y el ARN, generan una cantidad enorme de información en bruto y se requieren análisis bioinformáticos complejos para extraer información útil.

Para secuenciar un genoma, los millones de pares de bases del genoma se rompen en pequeños fragmentos de ADN que se secuencian repetidamente en paralelo, a continuación, se reconstruye o ensambla "*in silico*" el genoma de la persona comparándolo o cartografiándolo con uno o varios genomas de referencia a fin de identificar variaciones o variantes. El proceso de ensamblaje de la secuencia de un genoma a partir de millones de pequeños fragmentos de ADN está sujeto a error dependiendo de la eficiencia de los algoritmos estadísticos empleados para ello y de la calidad de los genomas de referencia empleados para guiar el ensamblaje. Lo que se conoce como "**sensibilidad y especificidad**" de la técnica depende, por una parte de la calidad y cantidad de fragmentos en bruto de ADN secuenciados que se producen y de los algoritmos empleados, si la calidad o cantidad es insuficiente (por ejemplo, debido a errores de secuenciación o vacíos [*gaps*]), es posible que haya variantes significativas que no sean

detectadas (sensibilidad baja) y por otra, de que haya variantes que puedan ser mal interpretadas como significativas (especificidad baja).

Aunque la precisión de la secuenciación para cada nucleótido individual es alta, dado el elevado número de nucleótidos en el genoma, si solo se secuencian una vez puede haber un número significativo de errores, para aumentar la precisión una región del genoma se secuencian múltiples veces, unas 30 para el diagnóstico de enfermedades raras y unas 100 en caso de tumores, el número de veces se denomina profundidad de lectura. Algunos tipos de variaciones genómicas causantes de enfermedad, como inserciones, deleciones e inversiones son particularmente difíciles de manejar debido a los desafíos que plantea su correcta colocación sobre el mapa del genoma de referencia (por ejemplo, no hay una única forma de posicionarlas). Las pruebas genéticas buscan alcanzar una sensibilidad y especificidad analíticas del 99% y en algunos estudios las NGS ya están alcanzando un estándar del 99,5%⁽⁸⁾, esta cifra puede parecer muy satisfactoria pero dados los miles o incluso millones de variaciones que podemos encontrar en el genoma este 0,5% que resta es importante. La precisión de una prueba para identificar variaciones específicas en una localización concreta del genoma es la “**validez analítica**”. Las NGS están todavía en fase de traslación y es clave que se consigan producir resultados suficientemente precisos y reproducibles que den el soporte adecuado a las decisiones clínicas⁽⁹⁾.

Las variaciones que se observan respecto al genoma de referencia se investigan obteniendo información de diversas fuentes, incluyendo bases de datos que relacionan variantes genéticas con determinados caracteres o condiciones de salud. Hay bases de datos de acceso público que recogen y comparten casos asociando una presentación clínica o enfermedad con una variante genética específica. Según la información que tenemos las variantes se clasifican en:

patogénicas, cuando hay una evidencia importante de que una variación genética se asocia a una enfermedad; **probablemente patogénicas** cuando la evidencia del impacto de la variante es menos sólida; **de significado incierto (VOUS)** cuando la evidencia no es clara; **probablemente benignas** y **benignas** también en función de las evidencias. La interpretación de variantes sin función conocida requiere una predicción precisa de su posible efecto mediante el software adecuado, no obstante, las predicciones pueden no reflejar la consecuencia biológica real, además, la misma variante puede tener predicciones de consecuencias diferentes según la base de datos consultada. Para demostrar que una variante es patogénica muchas veces hay que recurrir a estudios funcionales en células o tejidos del paciente, en líneas celulares en las que se haya generado la mutación o en modelos de otros organismos como las levaduras.

La interpretación de variantes es particularmente desafiante cuando la persona secuenciada es un recién nacido que no presenta signos o síntomas patológicos. En el ámbito del cribado neonatal no se deben comunicar las VOUS, sino únicamente las variantes patogénicas y las probablemente patogénicas, la comunicación de estas últimas no está exenta de controversia, pero la clasificación de variantes puede cambiar con el tiempo, y la mayoría de reclasificaciones en la base de datos ClinVar (el recurso de acceso libre del *National Center for Biotechnology Information* de la *US National Library of Medicine*, que archiva la relación entre variantes genómicas y la salud humana), van de probablemente patogénicas a patogénicas, de ahí la conveniencia de revelarlas. La evolución de la evidencia establece la necesidad de revisar periódicamente las variantes informadas en bases de datos actualizadas y establecer un flujo de trabajo para identificar proactivamente cambios en las categorías de variantes y comunicarlos, al mismo tiempo realizar la reevaluación clínica periódica en

los niños a riesgo y de las estrategias que apoyan la implementación del cribado basado en el subsiguiente manejo clínico^(9,10,11).

Por otra parte, es necesario el conocimiento de las variaciones o background (fondo genético) de la población para filtrar o descartar variantes comunes o polimorfismos. En las bases de datos empleadas para la interpretación de variantes están sobre representados los individuos caucásicos, por lo que no reflejan la diversidad étnica. El hecho de comparar con un genoma étnicamente emparejado o no, se estima que representa un 38% de reducción en la proporción de errores⁽¹²⁾. Por esto son muy importantes iniciativas como las del Centro de Investigación en Red de Enfermedades Raras (CIBERER) (www.ciberer.es), un consorcio dependiente del Instituto de Salud Carlos III a través del Ministerio de Ciencia e Innovación, creado para servir de referencia, coordinar y potenciar la investigación sobre las enfermedades raras en España, ya que diversos grupos de esta red han participado en la creación de la primera base de datos de la población española. La plataforma, denominada *Collaborative Spanish Variability Server* (CSVS) recoge un total de 2.027 genomas y exomas de individuos españoles no emparentados⁽¹³⁾. El grado en que una variante concreta se asocia a una enfermedad o carácter se conoce como “*validez clínica*”.

El genoma humano contiene unos 22.000 genes o secuencias de ADN, cada gen está dividido en exones e intrones, estos últimos son eliminados durante el proceso de transcripción por lo que solo se traducen a proteínas los exones. En total son unos 180.000 exones, que equivalen a unos 30 millones de pares de bases y representan solo alrededor del 2% del genoma completo. Podemos secuenciar por tanto el **genoma completo** (WGS o *whole genome sequencing*) que incluye la parte codificante del ADN y la no codificante, o podemos secuenciar solo la parte codificante del genoma, es decir

el exoma (WES o *whole exome sequencing*). Una vez secuenciado es posible capturar un **panel seleccionado de genes o de exomas diana** (*targeted sequencing* o secuenciación dirigida o diana) o incluso **limitar la búsqueda a un panel de variantes patogénicas conocidas**.

Para la secuenciación del exoma solo se capturan los exomas y las zonas intrónicas flanqueantes, por lo que no se pueden detectar las mutaciones intrónicas profundas. Son, así mismo, difíciles de detectar algunos tipos de inserciones y deleciones y no se pueden detectar las variaciones estructurales ni las repeticiones de nucleótidos. Cuando se secuencia un exoma se pueden encontrar alrededor de unas 60.000-70.000 variaciones, pero cuando se secuencia el genoma pueden ser millones. Cuanto mayor es la cantidad de información obtenida mayor es la probabilidad de encontrar los llamados hallazgos incidentales o no solicitados, que no están relacionados con la indicación que motivó la secuenciación, y que plantean cuestiones técnicas y éticas adicionales sobre la oportunidad de su revelación en función de la solidez científica del hallazgo y de la utilidad para el paciente^(14,15,16,17).

La secuenciación dirigida de paneles de genes diana o de exomas filtrando los genes de interés, es un método muy útil cuando tenemos información sobre los síntomas clínicos y buenas hipótesis diagnósticas. Podría también ser un método para buscar enfermedades diana concretas tratables en un programa de cribado neonatal.

Hay muchos factores que contribuyen a las incertidumbres que acompañan a la interpretación de los resultados de la secuenciación, errores en la propia secuenciación, anotaciones erróneas en las bases de datos, precisión limitada en las herramientas de predicción *in silico* para variantes con error de sentido y factores asociados a la propia complejidad genética. A menudo, es el análisis bioinformático el que nos marca la diferencia entre unos estudios y

otros y así puede darse el caso de que con una misma secuenciación, dependiendo de las herramientas bioinformáticas se identifiquen o no las mutaciones, por eso es importante conservar los datos primarios de la secuenciación para proceder al reanálisis bioinformático si fuera necesario. En el diagnóstico de enfermedades raras, tanto mediante secuenciación genómica como exómica, es frecuente que para la interpretación clínica se necesite secuenciar lo que se conoce como un “trío”, es decir el niño y los dos progenitores, lo cual es impracticable en un programa de cribado neonatal.

Muchos trastornos/enfermedades presentan una expresividad variable, lo cual significa que personas con el mismo trastorno pueden tener presentación clínica y síntomas muy diferentes. En otros casos puede que la enfermedad nunca se desarrolle, situación que se describe como penetrancia reducida o incompleta. Faltaría añadir a esta complejidad las posibles interacciones entre genes y la influencia de factores epigenéticos e interacciones medioambientales.

El paso último de la secuenciación es cuando se comunican los resultados, entendiendo como tales, la secuencia interpretada. El grado en que un diagnóstico molecular, es decir la comprensión de la base genética del trastorno, es informativo para el manejo clínico del paciente y se traduce en cambios sobre su salud se denomina “**utilidad clínica**”. Los resultados clínicamente accionables o con utilidad clínica, en un sentido estricto son los resultados de una prueba que son informativos para el manejo clínico de los pacientes y se traducen en cambios sobre su salud, incluyendo la maximización de beneficios mientras que al mismo tiempo se minimizan daños. Este equilibrio beneficios/riesgos depende de la eficacia de las intervenciones que siguen a la obtención del resultado y su subsiguiente aplicación⁽⁶⁾. La utilidad clínica depende también de la prevalencia de la enfermedad entre la población sometida a la prueba y puede

variar si se usa para un cribado de población general (prevalencia baja) o en personas con síntomas clínico o con una historia familiar con personas afectadas (prevalencia alta). El diagnóstico molecular es más probable que sea clínicamente útil para genotipos altamente penetrantes o genes con una elevada expresividad para un genotipo dado, pero en población asintomática su utilidad es muy limitada.

Las NGS, combinadas con otras tecnologías ómicas de alto rendimiento, epigenómica, transcriptómica, proteómica y metabolómica, contribuirán a descubrir las funciones de nuevos genes, servirán para identificar nuevas enfermedades y desarrollar nuevas aproximaciones terapéuticas, contribuirán al conocimiento de las enfermedades genéticas multifactoriales y serán muy relevantes para el mejor conocimiento de la genómica del cáncer y su tratamiento y de las enfermedades infecciosas. Pero tanto, en investigación básica como traslacional será imprescindible una aproximación interdisciplinar con la participación de clínicos, biólogos, bioquímicos y bioinformáticos.

LOS COSTOS REALES DE LA SECUENCIACIÓN GENÓMICA

En la introducción se ha comentado el espectacular descenso de los costos de la secuenciación genómica, que ya han roto la barrera de los 1000 dólares por genoma, cabe resaltar que hay centros españoles de excelencia como el CNAG-CRG (Centro Nacional de Secuenciación Genómica-Centro de Regulación Genómica) que realizan la secuenciación del genoma por 668 euros y la del exoma por 319 euros⁽¹⁸⁾.

El gigante de la secuenciación Illumina predijo en enero del 2017 que entre tres y diez años llegaríamos a los 100 dólares por genoma y a secuenciarlo en 1 hora⁽¹⁹⁾. Obtener el borrador de la secuencia en 2011 con el método de

Sanger requirió 11 años de trabajo y ahora ya se secuencian un genoma en 24-48 horas. Ha emergido incluso una tercera generación de tecnologías secuenciadoras que no requieren los pasos iniciales de amplificación del ADN, permitiendo la secuenciación de molécula única con la posibilidad de secuenciar hebras de ADN mucho más largas y hacerlo a tiempo real. *Oxford Nanopore Technologies* de Reino Unido ofrece su dispositivo portátil MinION que puede secuenciar ADN y ARN en tiempo real y fuera del laboratorio. Para una revisión de la evolución de las NGS se recomienda la lectura del siguiente artículo de G Ligtbody *et al*: *Review of applications of high-throughput sequencing in personalized medicine: barriers and facilitators of future progress in research and clinical application. Briefings in Bioinformatics*⁽²⁰⁾.

Pero el costo de secuenciación suele reflejar únicamente los de los componentes del equipo y los consumibles y no incluye los costos completos del proceso de secuenciación que incluyen el proceso de consentimiento informado, el procesado de las muestras, el de los datos bioinformáticos y su análisis, la interpretación de la secuencia, la verificación de variantes, la realización del informe, el almacenamiento de datos y el asesoramiento genético pre y post prueba. Las cifras que siguen tienen por objetivo matizar el concepto de que secuenciar un genoma es barato. Para dimensionar los costos reales es interesante el estudio realizado en Bélgica acerca del uso de la secuenciación genómica en la práctica clínica⁽²¹⁾ que explora los resultados obtenidos en cuatro estudios realizados en Reino Unido, Canadá, Alemania, y EE.UU. No todos incluyen los mismos conceptos, en general se asume una amortización del equipo en 5 años y los costos más elevados son siempre debidos al equipo y consumibles, las cifras oscilan entre 3.858 euros y 5.519 euros por muestra con el secuenciador Hi Seq 2500; 2.851 dólares con el Hi SeqX; 1.699 euros con el Hi Seq X Five y 1.411 euros con el Hi Seq X Ten.

Estos cálculos se refieren a actividad diagnóstica en pacientes con signos clínicos y secuenciando el genoma completo, sin embargo el cribado neonatal se dirige a población asintomática y por lo tanto el abordaje, como veremos más adelante, habría que delimitarlo y diseñarlo específicamente. Secuenciar exomas es obviamente menos costoso y secuenciar paneles de genes o exomas diana también, por ejemplo, en la referencia⁽²²⁾ se dan unos costos de 1.669 dólares por genoma completo, 792 dólares por exoma completo y 333 dólares por paneles de genes diana. En la referencia⁽²³⁾ se da un coste medio por exoma de 1.354 libras esterlinas. Si tomamos el valor más bajo de 1.411 euros por genoma los 400.000 recién nacidos anualmente en España costarían alrededor de 560 millones de euros, mientras que la del exoma supondría unos 317 millones.

Con los conocimientos actuales, como veremos más adelante, en un programa de cribado neonatal de salud pública la secuenciación dirigida a genes o exomas diana sería más apropiada, con lo que los costos, dependiendo de la interpretación y el volumen de datos, podrían disminuir apreciablemente respecto a la secuenciación del genoma completo. Así en el estudio realizado en Reino Unido que se comentará más adelante, mediante un cribado dirigido a la secuenciación exómica de únicamente 5 genes, obtienen unos costos/muestra de 71,14 libras esterlinas, si bien el cribado para las mismas enfermedades con los métodos actuales solo cuesta 25 libras esterlinas⁽²⁴⁾.

Será necesario emprender los estudios apropiados para obtener un cuadro realista de todos los costos y especialmente del coste-efectividad, incluyendo además de los costes analíticos y de almacenamiento y protección de datos, los de infraestructura y recursos humanos para garantizar la monitorización, educación y comunicación, asesoramiento genético, intervenciones, costos de vidas salvadas y años de

calidad de vida ganados así como los costes de oportunidad. La elevada heterogeneidad de los trastornos detectados por los programas actuales y la falta de evidencia científica sólida y a largo plazo sobre su historia natural y la efectividad de los tratamientos disponibles plantean una serie de dificultades metodológicas que limitan la aplicabilidad de los métodos clásicos de evaluación económica estándar. Por otro lado, la medición de los años de vida ajustados por calidad en población pediátrica y la falta de datos epidemiológicos, plantean desafíos metodológicos en un análisis de evaluación económica⁽²⁵⁾. Estas dificultades estarán también presentes en la evaluación del cribado neonatal genómico posiblemente incrementadas por las nuevas incógnitas que plantea.

EL DESAFÍO DE LA GESTIÓN, ALMACENAMIENTO Y PROTECCIÓN DEL DILUVIO DE DATOS QUE GENERAN LAS NGS

Las tecnologías de secuenciación avanzan tan rápidamente, sus costos descienden a tal velocidad y producen tal cantidad de datos, que el mayor desafío de la ciencia genómica para el siglo XXI es el desarrollo de una vasta estructura computacional y las herramientas de software para analizarlos, almacenarlos, acceder a ellos y compartirlos con las debidas garantías de seguridad, ya que los centros de secuenciación genómica serán algunos de los mayores usuarios de almacenamiento de datos. Como documento de consulta acerca de la gestión de datos genómicos con finalidad clínica y de investigación se recomienda la lectura del documento *Gestión de datos genómicos con finalidad clínica y de investigación*, del grupo de trabajo en gestión de datos genómicos del Instituto Roche 2015⁽²⁶⁾.

La secuenciación de un genoma genera aproximadamente 100 gigabytes (GB) de datos. La primitiva idea de secuenciar el recién nacido e

incorporarlos datos crudos de la secuencia a la historia clínica es poco factible porque las necesidades de memoria y velocidad de procesamiento son muy grandes para los servidores encargados hoy en día de la gestión de las historias clínicas de un centro. Es en cambio factible almacenar en la historia los informes de laboratorio que pueden contener la lista de variantes y su interpretación y los informes clínicos y de consejo genético.

La capacidad de computación necesaria para convertir las lecturas crudas al formato adecuado y la identificación posterior de variantes requiere servidores especializados con gran capacidad de computación y suficiente capacidad de almacenamiento dependiendo de la periodicidad de llegada de los datos crudos y el tiempo en que necesitamos dar el diagnóstico. La capacidad de computación va a determinar el tiempo requerido para poder concluir el análisis genómico y emitir el informe.

Dependiendo de estas necesidades, el almacenamiento y procesado pueden hacerse en servidores gestionados por los servicios de informática de la Institución, en equipos de terceros (*cloud computing*) o mediante modelos mixtos local-nube. También es posible externalizar el análisis computacional a un proveedor de servicios y almacenar los datos en un servidor institucional o archivar los datos en la nube con modelos de procesamiento remoto. Una opción interesante sería la de desarrollar infraestructuras de *cloud computing* públicas dentro del Sistema Nacional de Salud.

Si bien los costos de secuenciación han descendido espectacularmente, no lo han hecho al mismo ritmo los costos de los equipos informáticos y del almacenamiento de datos. En 2015 se estimaba el negocio de la genómica en la nube entre 100 y 300 millones de dólares con unas expectativas de crecimiento a 1.000 millones en 2018 y unos beneficios anuales

entre 50.000 y 70.000 millones⁽²⁷⁾. Algunos ejemplos de proveedores comerciales de nubes son *Amazon Web Service*, *Google Cloud Platform*, *Microsoft Azure* e *IBM Cloud*. Con unos costes de almacenamiento de 1GB de 0,015 dólares al mes⁽²⁸⁾, guardar un año los 400.000 genomas de los recién nacidos costaría 7,2 millones de dólares.

El almacenamiento de la información genética plantea cuestiones que van desde la gobernanza y la protección de la privacidad, a la estabilidad y accesibilidad de los datos. No hay almacenamiento sin riesgo potencial para la privacidad y confidencialidad y conforme se van acumulando datos genómicos en bases públicas aumenta la probabilidad de reidentificar una secuencia cruzando datos. Para las compañías DTC (de venta directa de pruebas genéticas a los consumidores) que ofrecen pruebas de antepasados y que almacenan millones de genomas de americanos con antepasados europeos, se ha estimado que un 60% de ellos podrían ser identificados partiendo de bases de datos de acceso público⁽²⁹⁾.

La seguridad de la e-medicina en la nube plantea desafíos importantes⁽³⁰⁾ y la protección mediante encriptación o *blockchain* se hace cada vez más necesaria. Por otra parte, irán emergiendo nuevas tecnologías secuenciadoras que proporcionen datos mejores y más precisos a costes cada vez más bajos, por lo que es posible que la primitiva idea de secuenciar el genoma del recién nacido y guardar la secuencia a lo largo de su vida para ir la reinterpretando quizás no tenga mucho sentido y sea más eficiente y menos costoso resecuenciar a la persona cuando sea relevante. En el contexto de los menores, el respeto por su privacidad, derecho a no saber y autonomía para dar su consentimiento cuando se alcance la edad legal sugieren además que es

prematureo almacenar la secuencia para futuros interrogantes⁽³¹⁾.

De hecho, hay una corriente que propone que una vez emitido el informe se destruya la secuencia en bruto evitando los costos y problemas del almacenamiento. Aunque esta opción limitaría la capacidad de investigación, tan necesaria para avanzar en la comprensión de la influencia de las variantes en la patología, que requiere el manejo de millones de datos genómicos.

La información que generaría secuencia el genoma de los 400.000 niños que nacen aproximadamente cada año en España sería de 4 millones de GB es decir alrededor de 40 petabytes (una magnitud de 16 cifras) y en poco tiempo llegaríamos a movernos en el ámbito de los hexabytes (19 cifras).

Aunque la secuenciación del genoma completo, como veremos más adelante, es una opción hoy en día muy cuestionada en el contexto del cribado neonatal, sí que emerge como una posibilidad interesante la secuenciación dirigida a paneles de genes o exomas para detectar enfermedades prevenibles o tratables en la infancia. Este abordaje generaría menos datos genómicos, pero requeriría igualmente una importante estructura computacional y de almacenamiento con unos costos también elevados y la correspondiente consideración de los aspectos éticos y legales asociados al manejo de los datos genómicos, cuyos puntos críticos conciernen al control de datos, la seguridad, la confidencialidad, la localización y la transferencia de datos, responsabilidad y protección; garantizando así mismo una conducta ética y el respeto para los participantes en la investigación en el caso de que los datos sean usados para el progreso de las ciencias biomédicas⁽³²⁾.

EL USO DE LAS NGS EN EL RECIÉN NACIDO ENFERMO CON FINALIDADES DIAGNÓSTICAS Y EN EL RECIÉN NACIDO SANO CON FINALIDADES DE SALUD PÚBLICA (CRIBADO NEONATAL). POTENCIAL Y LIMITACIONES

En 2013 el *National Institute of Child Health and Human Development* (NICHD) y el *National Human Genome Research Institute* (NHGRI) de EE.UU. patrocinaron el consorcio *Newborn Sequencing in Genomic Medicine and Public Health* (NSIGHT) que subvencionó con un total de 25 millones de dólares, cuatro proyectos de investigación acerca del uso de la secuenciación genómica en el recién nacido (tabla 1). Cada proyecto debía tener tres componentes: adquisición y análisis de datos genómicos en el periodo neonatal; investigación clínica para avanzar en la comprensión de los trastornos detectables mediante análisis del ADN y el estudio de las implicaciones éticas, legales y sociales del posible uso de la secuenciación genómica en recién nacidos. El objetivo general era aumentar la comprensión de cómo la secuenciación genómica podría mejorar tanto los cuidados de los recién nacidos enfermos como de los sanos en lo que respecta a las enfermedades genéticas. Los proyectos fueron revisados por el *NSIGHT Ethics and Political Advisory Board*. En el trabajo de Berg JS *et al*⁽³³⁾ se comparan los 4 proyectos financiados por el consorcio NSIGHT^(8,33,34,35,36,37,38,39,40). Tres de dichos proyectos incluyen recién nacidos sanos y enfermos, Nexus, BabySeq y NBSeq, este último emplea muestras de sangre seca remanentes de los programas de cribado neonatal almacenadas en el Biobanco de California y realizan la secuenciación en dos días. Un cuarto proyecto está dirigido a niños ingresados en la unidad de cuidados intensivos, las características de los cuatro proyectos se describen

en la tabla 1. Todos ellos se han realizado con asesoramiento genético pre y post prueba, consentimiento informado y revisión ética.

Antes de comentar los resultados del NSIGHT es preciso aclarar que en estos proyectos la noción habitual de enfermedades tratables, se sustituye por la más recientemente propuesta de enfermedades/genes medicamente accionables, noción que debe mucho a los principios generales del cribado propuestos por Wilson y Jungner en 1968⁽⁴¹⁾ basados en la gravedad e historia natural del trastorno, la capacidad de detectarlo en una etapa presintomática usando una prueba aceptable y la disponibilidad de un tratamiento así mismo aceptable. El término gen medicamente accionable se refiere típicamente a genes que cumplen los siguientes criterios:

- a) Las variantes causantes de enfermedad son suficientemente comprendidas.
- b) Es suficientemente probable que estas variantes causen enfermedad (son altamente penetrantes).
- c) Los efectos de la enfermedad son deletéreos.
- d) Hay intervenciones médicas, medioambientales o conductuales suficientemente eficaces y aceptables para prevenir o mejorar la enfermedad.

Encontrar métodos para evaluar la accionabilidad clínica es también muy relevante para decidir qué resultados incidentales y secundarios generados por las NGS deben revelarse⁽⁴²⁾. En los proyectos financiados por el NSIGHT se establecieron criterios de accionabilidad para decidir los genes asociados a enfermedad cuyos resultados se comunicarían. El proyecto BabySeq revisó 884 genes cuyas variantes patogénicas podían causar enfermedad en la infancia y los clasificó en tres categorías:

Tabla 1
Los cuatro proyectos financiados por el consorcio
Newborn Sequencing in Genomic Medicine and Public Health (NSIGHT).

El objetivo general era mejorar la comprensión de cómo la secuenciación genómica podría mejorar tanto los cuidados de los recién nacidos enfermos como de los sanos en lo que respecta a las enfermedades genéticas.

Universidad de North Caroline, Chapel Hill

Proyecto NEXUS. Secuenciación del exoma de recién nacidos sanos versus afectados con las enfermedades metabólicas hereditarias e hipoacusia, incluidas en los actuales programas de cribado neonatal. Reclutan 106 participantes y se hace secuenciación exómica de 466 genes seleccionados por su accionabilidad clínica. A un subgrupo de 65 padres se les propone elegir sobre recibir información acerca de enfermedades infantiles con poca o ninguna accionabilidad clínica, enfermedades de inicio en el adulto clínicamente accionables y el estado de portador para enfermedades autosómico recesivas, el 72,3% responde afirmativamente a las tres opciones. Definen como resultado positivo la observación de alguna variante patogénica o probablemente patogénica en alguno de los 466 genes, y como negativo la ausencia de variantes patogénicas o probablemente patogénicas. Se trabaja a ciegas y sin analizar “tríos” (niño+progenitores) ya que esto último sería impracticable en un programa de cribado neonatal. Confirman el diagnóstico en 15 de los 17 niños que tenían alguna enfermedad metabólica hereditaria y en 5 de los 28 con pérdida auditiva y encuentran 106 hallazgos incidentales. Encuentran hallazgos clínicamente accionables en 4 niños que no habrían sido detectados con el cribado estándar⁽³⁴⁾.

Brigham & Women’s Hospital / Harvard, Boston & Baylor College, Houston

Proyecto BabySeq. Es una investigación clínica aleatorizada que explora el impacto de la secuenciación exómica en dos cohortes, recién nacidos sanos y enfermos y evalúa las consecuencias para los niños, familia y clínicos. En cada cohorte los participantes se distribuyen aleatoriamente en dos grupos, los que recibirán únicamente cuidados estándar y los que además reciben la secuenciación genómica. Se recogen muestras de los recién nacidos y de los padres^(35,36). Se revisan 1.514 genes acerca de su asociación con enfermedades, edad de inicio, penetrancia y patrón de herencia para decidir qué resultados se retornarán. Se seleccionan finalmente 954 genes (se incluyen genes con fuerte evidencia de causar trastornos altamente penetrantes de inicio infantil, genes con evidencia y/o penetrancia moderada asociados a una enfermedad para la que una actuación en la infancia puede prevenir una enfermedad devastadora más tarde, genes con fuerte asociación farmacológica relevante para tratamientos pediátricos y estado de portador para cualquier gen que cumpla criterios⁽³⁷⁾). Solo se reportan variantes patogénicas o probablemente patogénicas en los niños sanos, en los enfermos se incluyen variantes de significado incierto para genes relevantes de acuerdo con la indicación clínica. En total se han reclutado 159 recién nacidos (127 sanos y 35 enfermos). Identifican enfermedades genéticas de inicio en la infancia en 9,4% de los recién nacidos, riesgo para enfermedad accionable de inicio en la edad adulta en un 3,5 % y una variante de portador heterocigoto en un 88%.

Universidad de California, San Francisco

Proyecto NB Seq. Secuenciación en dos días del ADN de muestras de sangre seca para mejorar y expandir el cribado neonatal. Capturan y analizan 78 genes implicados en enfermedades metabólicas hereditarias. Las muestras proceden del biobanco de California de muestras residuales de los programas de cribado neonatal. Realizan la secuenciación exómica de 1.190 muestras asociadas a datos clínicos de los recién nacidos, incluyendo una cohorte de 50 pacientes con inmunodeficiencias. Comparan los resultados obtenidos secuenciando el exoma con los obtenidos empleando la espectrometría de masas en tándem⁽³⁹⁾.

Children’s Mercy Hospital, Kansas City

Proyecto recién nacidos severamente enfermos. Estudian las implicaciones clínicas y sociales de los resultados de la secuenciación completa del genoma en dos días para acelerar el diagnóstico etiológico en niños críticamente enfermos. Los 65 participantes (edad ≤ 4 meses) fueron distribuidos de forma aleatoria en dos grupos, los que se estudiaban con las pruebas genéticas habituales y los que recibían la secuenciación genómica rápida de “tríos” (niño + progenitores). La sensibilidad y especificidad de la detección de variantes fue $>99,5\%$. Los niños que fueron secuenciados fueron diagnosticados en un promedio de 25 días versus los estudiados con las pruebas genéticas convencionales que fue de 130 días⁽⁸⁾. Un segundo ensayo NSIGHT 2 ha demostrado que no hay una diferencia importante en la tasa de diagnóstico entre secuenciación genómica y secuenciación exómica rápidas. Mientras que la secuenciación genómica es técnicamente más completa y precisa, los dos métodos son equivalentes en términos de tiempo en obtener el resultado 11,8 días y tasa de diagnóstico, un 20% de los niños estudiados⁽⁴⁰⁾.

- a) Evidencia fuerte o definitiva de causar un trastorno altamente penetrante.
- b) Evidencia moderada o penetrancia moderada, potencialmente accionable en la infancia.
- c) Evidencia insuficiente, penetrancia baja o moderada o de inicio en la edad adulta sin accionabilidad en la infancia^(37,38).
- iii) Enfermedad de inicio en la edad adulta con accionabilidad elevada.
- iv) Enfermedad de inicio en la edad adulta con accionabilidad baja o ninguna⁽³⁸⁾, en esta referencia se comparan los dos sistemas de puntuación y se pone de manifiesto que 244 genes/enfermedades coinciden en ser incluidas en la categoría I de Nexus y en la A de BabySeq, serían pues las más interesantes de cara a su inclusión en un programa de cribado neonatal.

El proyecto Nexus estableció un sistema semicuantitativo de puntuación basado en la edad de aparición de síntomas (infancia, edad adulta) y un nivel de accionabilidad según la información disponible (elevada, baja o ninguna) y clasificó las enfermedades en cuatro categorías:

- i) Enfermedad de inicio en la edad pediátrica con accionabilidad elevada.
- ii) Enfermedad de inicio pediátrico con accionabilidad baja o ninguna.

El paso siguiente sería aplicarles clasificaciones basadas en la evaluación de su validez clínica⁽⁴³⁾.

Las conclusiones alcanzadas con los 4 proyectos NSIGHT respecto a la adquisición y análisis de datos genómicos en el periodo neonatal y la investigación clínica para avanzar en la comprensión de los trastornos se han resumido en las tablas 2a y 2b. En la primera se exponen

Tabla 2a
Conclusiones alcanzadas con los cuatro proyectos financiados por el consorcio *Newborn Sequencing in Genomic Medicine and PublicHealth* (NSIGHT) en el CONTEXTO DIAGNÓSTICO.

Ver las conclusiones en el contexto salud publica en la tabla 2b^(8,34,35,36).

En el recién nacido severamente enfermo con trastornos genéticos, la secuenciación genómica usada precozmente acelera el diagnóstico definitivo, estudios preliminares demuestran incluso una potencial reducción de costos diagnósticos frente al uso de las técnicas genéticas habituales⁽⁸⁾.

Su implantación viene frenada por las dificultades en la interpretación de resultados ambiguos, elevado número de hallazgos secundarios, potencial ansiedad paterna, y factores económicos, incluyendo la resistencia a pagar por pruebas para las que todavía no hay datos definitivos de utilidad clínica y coste-efectividad⁽⁸⁾.

Se requiere un estudio en mayor profundidad de la respuesta de los padres y neonatólogos a la secuenciación genómica y el desarrollo de un marco riguroso para trasladar esta información a los planes de neonatología de precisión en las UCI de neonatos⁽⁸⁾.

Tabla 2b
Conclusiones alcanzadas con los cuatro proyectos financiados por el consorcio
Newborn Sequencing in Genomic Medicine and PublicHealth (NSIGHT)
en el CONTEXTO SALUD PÚBLICA (cribado neonatal)^(8,34,35,36,37,38,39).

A nivel poblacional es improbable que la secuenciación genómica pueda reemplazar al cribado neonatal actual basado en marcadores bioquímicos o audiométricos debido a la heterogeneidad etiológica y los desafíos de la interpretación de variantes⁽³³⁾.

Para algunas enfermedades metabólicas hereditarias la secuenciación exómica funciona tan bien como la detección mediante espectrometría de masas en tándem, pero para otras no, se ha de considerar gen por gen^(33,36,39).

Por el momento la secuenciación exómica es inadecuada como único método de cribado para enfermedades metabólicas hereditarias^(31,34,37).

Aunque asumimos que las enfermedades mendelianas serán mejor identificadas con la secuenciación genómica, está por demostrar y puede no ser cierto. Las enfermedades monogénicas están a menudo influenciadas por factores adicionales aún no identificados, genéticos y ambientales, mientras que la espectrometría de masas en tandem mide analitos que son muy cercanos al fenotipo^(35,37,39).

La secuenciación genómica puede funcionar sin embargo como una segunda prueba para aumentar la especificidad del MS/MS ayudando a distinguir los falsos positivos de los verdaderamente enfermos y ayudar al diagnóstico diferencial de perfiles bioquímicos inespecíficos. Podría mejorar el conocimiento de la relación fenotipo/genotipo y revelar posibles contribuciones genéticas a los falsos positivos^(33,35,39).

La falta de datos respecto a su performance analítica y clínica como prueba predictiva constituye un desafío significativo para el uso de la secuenciación en el cribado neonatal. Se sabe muy poco acerca del valor positivo o negativo predictivo en población asintomática cuya probabilidad de enfermedad es a priori muy pequeña⁽³³⁾.

Los algoritmos de selección de variantes maximizan la sensibilidad con lo que necesariamente sacrifican la especificidad llevando a un aumento de falsos positivos y posibles daños debido a intervenciones médicas innecesarias. Sería necesaria una estrategia que excluyese las variantes de significado incierto que por definición tienen poco valor predictivo. En una población con riesgo bajo como es la de la población sana de recién nacidos, la revelación de hallazgos genómicos con certeza baja dispararía los falsos positivos⁽³³⁾.

Históricamente los programas de cribado neonatal de trastornos prevenibles tienen baja tolerancia para los falsos negativos. Con la secuenciación genómica sería necesario desplazar el paradigma del cribado neonatal de encontrar todos los individuos afectados por la detección de una proporción óptima de casos para un número más elevado de enfermedades potencialmente tratables^(33,34).

La práctica de retornar los resultados de heterocigoto portador podría ser insostenible dado que cada individuo es posiblemente portador para varias enfermedades raras^(33,34).

En el contexto de salud pública hay que seleccionar bien que enfermedades cribar considerando factores como la edad de inicio, severidad, penetrancia, posibilidades de tratamiento y pruebas de confirmación. Algunas enfermedades genéticas podrían satisfacer los criterios de Wilson & Jungner⁽⁴²⁾, pero muchas no^(33,34).

La secuenciación genómica o exómica podría ser usada como prueba de cribado para enfermedades raras para las que no se ha encontrado marcador bioquímico y dado que puede identificar virtualmente cualquier enfermedad de causa genética, alguna forma de secuenciación genómica es inevitable; a corto plazo la secuenciación de genes ya asociados a las enfermedades cribadas actualmente podría mejorar su sensibilidad y especificidad^(33,34).

Expandir los paneles actuales para incluir otras enfermedades accionables detectables únicamente mediante secuenciación, podría potenciar los beneficios de salud pública de los programas actuales siempre y cuando se superen las dificultades interpretativas para equilibrar la proporción de casos detectados versus los falsos positivos^(33,34).

Los desafíos en relación a los costos, implementación y consideraciones éticas para llevar a cabo la secuenciación a gran escala en niños sanos limitan su aplicación a toda la población y ensombrecen la realización de la promesa del Proyecto Genoma Humano^(33,34).

las conclusiones en el contexto diagnóstico y en la segunda las conclusiones en el contexto salud pública (cribado neonatal). Las conclusiones obtenidas respecto las implicaciones éticas, legales y sociales se comentarán en la Parte II de este artículo.

Otro proyecto que vale la pena mencionar es el llevado a cabo por el *United Kingdom National Health Service*⁽²⁴⁾ para explorar el potencial diagnóstico y pronóstico de las NGS, porque no solo valida la tecnología para ADN extraído de muestras de sangre seca comparándola con el extraído de muestras de sangre venosa, sino que automatiza la secuenciación y el proceso bioinformático para procesar 1.000 muestras a la semana y conseguir que solo transcurran cuatro días entre el procesado de la muestra y entrega del resultado. Es decir, reproduce las condiciones reales si se aplicase a un programa de cribado neonatal. El protocolo se ha realizado con muestras de adultos voluntarios sanos y fue aprobado por el *London-Brent Research Ethics Committee*. Se realiza una secuenciación dirigida del exoma de 5 genes diana, concluyen que como primera prueba de cribado la técnica no es coste-efectiva, pero que como segunda prueba de cribado sería útil para clasificar casos ambiguos de los cribados actuales, mejorar la utilidad pronóstica de sus resultados y contribuir a la medicina de precisión seleccionando los tratamientos óptimos. La utilidad de las NGS como segunda prueba para la mejor clasificación y manejo de los casos positivos de los cribados actuales se pone también de manifiesto en las conclusiones de los proyectos NSIGHT (tabla 2), y también hay estudios realizados en España que lo apoyan⁽⁴⁴⁾.

En el momento de escribir este artículo se ha puesto en marcha en España el proyecto piloto “Gen Natal”, para analizar cómo implementar la secuenciación genómica en medicina neonatal y salud pública, financiado por la Fundación Ramón Areces con 180.000 euros en el que

participan unidades del CIBERER con un planteamiento muy próximo al de los proyectos NSIGHT de EE.UU.⁽⁴⁵⁾.

Dadas las conclusiones de los proyectos NSIGHT, el hecho de que las NGS están todavía en fase de traslación, siendo clave que se puedan producir resultados suficientemente precisos y reproducibles para dar soporte a la toma de decisiones clínicas⁽⁹⁾ y las consideraciones generales sobre cribados basados en el ADN⁽¹¹⁾, podemos sin embargo anticipar que **si se introdujese la secuenciación genómica en el programa de cribado neonatal tendría un gran impacto en la organización y renovación de los centros de cribado**, ya que deberían garantizarse entre otros aspectos la calidad, precisión y reproducibilidad de resultados intra e inter centros y abordar aspectos como: la ampliación de personal con competencias en genética molecular y bioinformática; tomar decisiones sobre los equipos secuenciadores (compra, externalización de la secuenciación y recepción online de la secuencia para su interpretación, profundidad de lectura); manejo de las variantes (filtrado y anotación, líneas guía para la interpretación de variantes, fiabilidad de las bases de datos públicos, filtrado de hallazgos incidentales para minimizar su impacto); necesidad de revisar periódicamente las variantes informadas en bases de datos actualizadas y establecer un flujo de trabajo para identificar proactivamente cambios en las categorías de variantes y comunicarlo; buenas prácticas de laboratorio (precisión y reproducibilidad, estandarización del análisis bioinformático y métodos estadísticos, materiales de referencia certificados); interpretación y formato del retorno de resultados, tiempo de entrega; herramientas computacionales, almacenamiento y manejo de la base de datos (tipo de servidor, nube, duración del almacenamiento, acceso, seguridad). La gran cantidad de datos que se generarían con 400.000 recién nacidos al año debería contemplar el papel de las herramientas de la inteligencia artificial en la mejora de la comprensión y

priorización de variantes y considerar el potencial del “*deep learning*”.

CONCLUSIONES

El conocimiento del genoma humano es aún incipiente, todavía no se conocen al completo las relaciones gen-enfermedad y el significado de muchas variaciones en la secuencia no es claro.

La expresión de muchas variantes genéticas está fuertemente influenciada por un complejo sistema de fuerzas genómicas, epigenéticas y medioambientales que son distintas para cada individuo y las predicciones en base únicamente a la secuencia pueden ser inciertas.

La interpretación de variantes es especialmente difícil si el individuo secuenciado no tiene síntomas clínicos como son la inmensa mayoría de recién nacidos. No debe olvidarse que el cribado neonatal, es un programa de cribado universal, dirigido a toda la población de recién nacidos, estando, de hecho, incluido en el programa para el niño sano.

Las tecnologías de secuenciación masiva han demostrado en los proyectos de investigación un potencial extraordinario como herramienta diagnóstica, pero conviene resaltar que todavía estamos en fase de traslación para su aplicación en la clínica y quedan unos cuantos desafíos a superar antes de que puedan revertir todo su potencial a los pacientes, los clínicos y la sociedad, la clave es si pueden producir resultados suficientemente precisos y reproducibles para dar soporte a la toma de decisiones clínicas.

La distancia entre lo que los resultados de la secuenciación pueden revelar y la clase de información que la mayoría de personas necesita para mejorar su salud, en combinación con las ampliamente publicitadas esperanzas acerca del poder revolucionario de la genómica, crea

un riesgo muy real de que pacientes, participantes en la investigación, profesionales de la salud, políticos y agencias reguladoras, tengan expectativas poco realistas de lo que la secuenciación puede alcanzar y escasa apreciación de sus inconvenientes.

La falta de datos respecto a la validez analítica y clínica de las NGS como prueba predictiva constituye un desafío significativo para el uso de la secuenciación en el cribado neonatal. Se sabe muy poco acerca del valor positivo o negativo predictivo en población asintomática cuya probabilidad de enfermedad es, a priori, muy pequeña.

Se espera que los datos genómicos puedan mejorar estrategias preventivas y personalizadas, pero también pueden obtenerse falsos negativos y aumento de los falsos positivos dependiendo de los paneles y las decisiones sobre variantes a comunicar, con el consiguiente sobrediagnóstico, sobretratamiento y pacientes en espera. Las evidencias robustas y la comprensión de la presencia de las variantes en la población son un prerrequisito para un cribado neonatal responsable y efectivo.

El cribado neonatal no debe usarse como un sustitutivo de unos buenos servicios de diagnóstico. Hay un consenso en que en el estado actual de conocimientos, ni la secuenciación genómica completa ni la de paneles de genes extensos debe ser implementada como cribado neonatal universal, y también en que la secuenciación genómica dirigida a enfermedades concretas en recién nacidos asintomáticos no debe substituir a los programas con los métodos de cribado actuales, sino que debe ser considerado como una ampliación para incluir enfermedades infantiles tratables que no puedan ser diagnosticadas con otros métodos.

La secuenciación genómica o exómica, podría ser usada como prueba de cribado para

enfermedades raras que sean detectables únicamente mediante secuenciación potenciando los beneficios de salud pública de los programas actuales, siempre y cuando se superen las dificultades interpretativas para equilibrar la proporción de casos detectados versus falsos positivos. Seleccionando bien las enfermedades, considerando factores como la edad de inicio, gravedad, penetrancia y posibilidades de tratamiento, algunas enfermedades podrían satisfacer los clásicos criterios de Wilson y Jungner⁽⁴¹⁾ para los cribados poblacionales, pero muchas no.

Pero incluso con estas premisas, la introducción de la secuenciación genómica dirigida en el programa de cribado neonatal tendría un gran impacto en la organización y renovación de los centros de cribado. Deberán emprenderse los estudios apropiados para obtener un cuadro realista de los costos y del coste efectividad y considerar los aspectos éticos sociales y legales.

La secuenciación genómica puede funcionar como una segunda prueba para los casos positivos detectados con espectrometría de masas en tándem, ayudando a distinguir los falsos positivos de los verdaderamente enfermos, mejorando el conocimiento de la relación fenotipo/genotipo y revelando posibles contribuciones genéticas a los falsos positivos.

La primitiva idea de secuenciar el recién nacido y guardar la secuencia a lo largo de su vida para ir la reinterpretando quizás no tenga mucho sentido en la actualidad, dado que seguirán emergiendo nuevas tecnologías secuenciadoras que proporcionen datos mejores y más precisos a costes cada vez más bajos, por lo que será más eficiente y menos costoso resecuenciar a la persona cuando sea relevante. En el contexto de los menores, el respeto por la privacidad, derecho a no saber y autonomía para dar su consentimiento cuando se alcance la edad legal sugieren además que es prematuro almacenar la secuencia para futuras interrogaciones.

En base a los conocimientos actuales, el potencial de producir resultados falsos positivos o resultados simplemente ambiguos o no interpretables sobrepasa los potenciales beneficios de la información probabilística, por lo que la secuenciación genómica no dirigida no debe ser recomendada en el ámbito del cribado neonatal, si bien también hay que admitir que la situación es fluida y la única forma de adquirir conocimiento es seguir investigando.

CONTRIBUCIÓN DE AUTORÍA

El autor para correspondencia asumió la preparación del borrador original y la formulación de la estructura del artículo. Todos los autores han contribuido por igual en la conceptualización, análisis formal y revisión crítica.

BIBLIOGRAFÍA

1. Department of Health. Our inheritance our future. Realizing the potential of genetics in the NHS. June 2003.
2. Francis S Collins. The language of life. The DNA and the revolution in personalized medicine. 2010.
3. Alkmaya FS. A genetic revolution in rare-disease medicine. *Nature* 2021; 590: 218-219.
4. Saunders CJ, *et al.* Rapid Whole Genome sequencing for genetic disease diagnostic in neonatal intensive care units. *2012 Science Translational Medicine*; 154(4): 154-155.
5. Johnston J, Lantos JD, Goldenberg A, Chen F, Parens E, Koenig BA, and members of the NSIGHT Ethics and Policy Advisory Board. Sequencing newborns: A call for nuanced use of genomic technologies. In *The ethics of sequencing newborns: Recommendations and Reflections*, especial report. *Hastings Center Report* 48, nº 4 (2018): S2+ [20 pages]. doi: 10.1002/hast.874
6. Grosse S.D, Kalman L, Khoury MJ. Evaluation of the Validity and utility of Genetic testing for Rare Diseases. En *Rare disease epidemiology*. En Posada de la Paz M,

- Groft SC (eds), Rare disease epidemiology. Advances in Experimental Medicine and Biology 686. Springer 2010. Pp 115-121.
7. ACCE Model Process for Evaluating Genetic Tests. <http://www.cdc.gov/genomics/gtesting/acce>
 8. Petrikin JE, Cakici JA, Clark MM *et al*. The NSIGHT 1-randomized controlled trial: rapid whole-genome sequencing for accelerated etiologic diagnosis in critically ill infants. *NPJ Genomic Medicine* 2018; 3:6; doi: 10.1038/541525-018-0045-8
 9. Liu Z, Zhu I, Roberts R, Tong W. Toward clinical implementation of next-generation sequencing-based genetic testing in rare diseases: where are we? *Trends in Genetics* 2019; 35 (11): 852-867.
 10. Holm IA, Yu TW, Joffe S. From sequence data to returnable results: Ethical issues in variant calling and interpretation. *Genet Test Mol Biomarkers* 2017, 21(3): 178-183.
 11. Murray M, Giovani MA, Doyle DL, Harrison SM *et al* and the American College of Medical Genetics and Genomics board of directors. DNA-based screening and population health: a point to consider statement for programs and sponsoring organizations from the American College of Medical Genetics and Genomics (ACMG). *Genetics in medicine* 2021. <https://doi.org/10.1038/s41436-020-01082-w>
 12. Luheshi L, Raza S. Clinical whole genome analysis: delivering right diagnosis. PHG foundation January 2014
 13. Peña-Chilet M, Roldán G Pérez Florido *et al*. CSVS; a crowdsourcing database of the Spanish population genetic variability. *Nucleic Acids Research* 2020. doi: 10.1093/nar/gkaa794
 14. Presidential Commission for the Study of Bioethical Issues. Anticipate and communicate. Ethical management of incidental and secondary findings in the Clinical, Research and Direct -to-Consumer Contexts. December 2013.
 15. PHG foundation. Making science work for health. Managing incidental and pertinent findings from whole genome sequencing in the 100.000 genomes project. A discussion paper from the PHG foundation. April 2013.
 16. Hehir-Kwa JY, Claustres M, Hastings RJ *et al*. Towards an european consensus for reporting incidental findings during clinical NGS testing. *Eur J Hum Genet.* 2015; 23: 1601-1606.
 17. Holm IA. Pediatric issues in return of results and incidental findings: weighing autonomy and best interest. *Genetic testing and molecular biomarkers* 2017; 21(3), 155-1587.
 18. CNAG-CRG (Centro Nacional de Secuenciación Genómica-Centro de Regulación Genómica). Newsletter 15 de enero de 2021.
 19. Fikes B. New machines can sequence human genome in one hour, Illumina announces. The San Diego Union Tribune. 2017 <http://www.sandiegouniontribune.com/bussines/biotech/sd-me-illumina-novaseq-20170109-story.html>
 20. Ligtbody G, Haberland V, Browne L *et al*. Review of applications of high-throughput sequencing in personalized medicine: barriers and facilitators of future progress in research and clinical application. *Briefings in Bioinformatics.* 2019; 20(5): 1795-1811.
 21. Hanket G, Vinck I, Thiry N. The use of whole genome sequencing in clinical practice: challenges and organizational considerations for Belgium. *Health Services Research (HSR) Brussels.* Belgian Health Knowledge Centre KCE Reports 300.D/2018//10.273/25.
 22. Van Ninwegen KJM, van Soes AA, Veltman JA *et al* Is the \$100 genome as near as we think? A cost analysis of next generation sequencing. *ClinChem* 2016; 62(11) 1458-1464.
 23. The budgeted impact and cost-effectiveness of introducing whole exome sequencing-based virtual gene panel test into routine clinical genetics. PHG Foundation 2017 ISBN 978-1-907198-25-0).
 24. Van Campen JC, Sollars ESA, Thomas RC *et al*. Next generation sequencing in newborn screening in the United

- Kingdom National Health Service. *Int J Neonatal Screen*. 2019; 5: 40. doi: 10.3399/ijns 5040040
25. Valcarcel Nazco C, García Pérez L, Linertová R, Castilla I, Vallejo Torres I, Ramos Goñi JM, Labrador Cañadas V, Couce ML, Espada Sáenz-Torres M, Dulin Iñiguez E, Posada M, Imaz Iglesias I, Serrano Aguilar P. Métodos para la evaluación económica de programas de cribado neonatal. *RevEsp de Salud Pública*. 2021; 95: 26 de enero e 202101009.
26. Grupo de trabajo en gestión de datos genómicos. Gestión de datos genómicos con finalidad clínica y de investigación. Instituto Roche 2015.
27. Amazon, Google race to get your DNA in the cloud. *Medscape*. Jun 05, 2015.
28. Krimm N, Hoffman N. Practical estimation of cloud storage costs for clinical genomic data. *Practical Laboratory Medicine* 2020; 21, e00168.
29. Erlich Y, Shor T, Peter II. Identity inference of genomic data using long range familial searches. *Science* 2018; 362:690-694.
30. Al-Issa Y, Ottom MA, Tamrawi A, eHealth Cloud Security Challenges: A survey. *Journal of Health care engineering* 2019: 1-17. Article ID 7516035. <https://doi.org/10.115572019/7516035>
31. Howard CH, Knoppers BM, Cornel MC, Clayton EW, Sénécal K, Borry P, endorsed by the European Society of Human Genetics; the P3G International Pediatric Platform; the Human Genome Organization; and the PHG Foundation. Whole-genome sequencing in newborn screening. A statement on the continued importance of targeted approaches in newborn screening programs. *Eur J Hum Genet* 2015; 23: 1593-1600.
32. Dove ES, Joly Y, Tassé AM, Public Population project in Genomics and society (P3G) International Steering Committee, International Cancer Genome Consortium (ICGC) Ethics and policy Committee and Knoppers BM. Genomic cloud computing: legal and ethical points to consider. *Eur J Hum Genet*. 2015; 23: 1271-1278.
33. Berg JS, Agrawal PB, Bayley DB *et al*. Newborn screening in genomic medicine and public health. *Pediatrics* 2017; 139(2): e20162252.
34. Roman TS, Crowley Sb, Rocha MI *et al*. Genomic sequencing for newborn screening. Results of the NC NEXUS project. *Med Rxiv preprint* doi. <https://doi.org/10/1101/2020.02.26.20024679>. Posted february 29, 2020.
35. Holm IA, Agrawal PB, Ceyhan-Birsoy O *et al*. The BabySeq Project: implementing genomic sequencing in newborns. *BMC pediatrics* 2018; 18: 225. <https://doi.org/10.1186/s12887-018-1200-1>
36. Ceyhan-Birsoy O, Murry JB, Machini K *et al*. Interpretation of genomic sequencing results in healthy and ill newborns: Results from de BabySeq project. *Am J Hum Genet* 2019; 104: 76-93.
37. Ceyhan-Birsoy JB; Machini K, Lebo MS *et al*. A curated gene list for reporting results of newborn genomic sequencing. *Genet Med* 2017; 19(7): 809-818.
38. Milko LV, O'Daniel S DeCristo DM *et al*. An age based framework for evaluating genome-scale sequencing results in newborn screening. *J Pediatr* 2019; 209: 68-76.
39. Adkiahari AN, Gallagher R; Wang Y *et al*. The role of exome sequencing in newborn screening for inborn errors of metabolism. *Nature Medicine* 2020; 26: 1392-1397.
40. Kingsmore SF, Cakici JA, Clark MM *et al*. A randomized controlled trial of the analytic and diagnostic performance of singleton and trio rapid genome and exome sequencing. *The Am J Hum Genet* 2019; 105: 719-733.
41. Wilson JM, Jungner G. Principles and practice of screening for disease. *Public health papers n° 34*. World Health Organization. Geneva. 1968.
42. Berg JS, Foreman AK, O'Daniel, JM, Booker JK, Boshe L, Carey T *et al*. A semiquantitative method for evaluating clinical actionability of incidental and secondary findings from genome scale sequencing. *Genet Med*. 2016; 18. A67-475.

43. Strande NT, Riggs Er, Buchanan AH, Ceyhan-Birsoy O, DiStefano m, Dwigt SS *et al*. Evaluating the clinical validity of gene-disease association. An evidence-based framework developed by the clinical genome resource. *Am J Hum Genet.* 2017; 100: 895-906.
44. Navarrete R, Leal F, Vega AI, Morais-López A, García-Silva MT *et al*. Value of genetic analysis for confirming in-born errors of metabolism detected by the Spanish neonatal screening program. *Eur J Hum Genet.* 2019; 27: 556-562.
45. Palau F. Medicina genómica y salud pública en el recién nacido: ampliación del cribado neonatal a otras patologías genéticas raras. En 50 años de cribado neonatal: como afrontamos el futuro. Coord Belén Pérez González. Fundación Ramón Areces 2021, págs. 77-78. (Accesible en www.fundacionareces.es).