



UNIVERSIDAD NACIONAL AUTÓNOMA DE MÉXICO

FACULTAD DE QUÍMICA

**DE LA SECUENCIACIÓN
DEL GENOMA
A LA MEDICINA DE PRECISIÓN**

**TRABAJO MONOGRÁFICO DE
ACTUALIZACIÓN**

Que para obtener el título de

Químico Farmacéutico Biólogo

P R E S E N T A

Mariana Calzada Mijangos

DIRECTOR(A) DE TESINA

Dra. Georgina Hernández Montes

Ciudad Universitaria, Cd. Mx., 2022



Universidad Nacional
Autónoma de México



UNAM – Dirección General de Bibliotecas
Tesis Digitales
Restricciones de uso

DERECHOS RESERVADOS ©
PROHIBIDA SU REPRODUCCIÓN TOTAL O PARCIAL

Todo el material contenido en esta tesis esta protegido por la Ley Federal del Derecho de Autor (LFDA) de los Estados Unidos Mexicanos (México).

El uso de imágenes, fragmentos de videos, y demás material que sea objeto de protección de los derechos de autor, será exclusivamente para fines educativos e informativos y deberá citar la fuente donde la obtuvo mencionando el autor o autores. Cualquier uso distinto como el lucro, reproducción, edición o modificación, será perseguido y sancionado por el respectivo titular de los Derechos de Autor.

Tabla de contenido

1	ANTECEDENTES	3
1.1	PRÁCTICA MÉDICA ACTUAL	3
2	IMPLEMENTACIÓN DE LA MEDICINA DE PRECISIÓN	6
3	DESARROLLO DE TECNOLOGÍAS DE ALTO RENDIMIENTO CON APLICACIÓN MÉDICA	12
4	INVESTIGACIÓN BÁSICA	13
4.1	GENÓMICA	14
4.2	TRANSCRIPTÓMICA	25
4.3	EPIGENÓMICA	29
4.4	FARMACOGENÓMICA	33
4.5	EXPOSÓMICA	35
4.6	METABOLÓMICA	39
4.7	PROTEÓMICA	45
4.8	INTEGRACIÓN DE DATOS OBTENIDOS MEDIANTE CIENCIA BÁSICA	47
5	INVESTIGACIÓN APLICADA	50
5.1	BIOMARCADORES	50
5.2	REPOSICIONAMIENTO DE FÁRMACOS	55
6	BIOBANCOS	57
7	HISTORIAS CLÍNICAS ELECTRÓNICAS	61
8	INFORMÁTICA BIOMÉDICA TRANSLACIONAL	68
8.1	INFRAESTRUCTURA	68
8.2	BASES DE DATOS	71
8.3	ALGORITMOS Y ANÁLISIS DE <i>BIG DATA</i>	71
8.4	SEGURIDAD INFORMÁTICA	74
9	LEGISLACIÓN	76
9.1	NORMAS, LEYES Y TRATADOS	76
9.2	BIOÉTICA	85
10	BENEFICIOS DE LA IMPLEMENTACIÓN	90
10.1	BENEFICIOS MÉDICOS	90
10.2	BENEFICIOS ECONÓMICOS	92
11	ESTADO DEL ARTE DE LA MEDICINA DE PRECISIÓN A NIVEL GLOBAL	98
12	CONCLUSIONES	104
13	BIBLIOGRAFÍA	108

1 Antecedentes

1.1 Práctica médica actual

Desde su surgimiento, la medicina es una ciencia que busca ser precisa y se encuentra en ese proceso. En la práctica médica actual, el médico evalúa los síntomas del paciente y con base en su diagnóstico decide qué pruebas deben realizarse para reunir más información. De la misma manera, será el responsable de establecer y determinar la causa de los síntomas y el pronóstico del paciente. Sin embargo, para que este sistema sea efectivo y preciso, el médico necesitaría evaluar toda la información actualizada sobre la historia natural de la enfermedad e incorporar lo que conoce acerca del paciente en el proceso de la toma de decisiones.

Por otra parte, actualmente la clasificación de la mayoría de las patologías está basada únicamente en la información clínica de la enfermedad sustentada en sus signos y síntomas. Una de las desventajas de este sistema es clasificar como una misma enfermedad cuando se trata de subtipos de enfermedades con causas moleculares distintas, y además se han clasificado como diferentes enfermedades a patologías que comparten una causa molecular común. Esta taxonomía usualmente se refiere a la Clasificación Internacional de Enfermedades (ICD, *International Classification Disease* por sus siglas en inglés), un sistema establecido hace más de 100 años por la Organización Mundial de la Salud (OMS) y que sigue siendo usada por los médicos para realizar diagnósticos estandarizados y que en ocasiones son poco precisos. Considerando que en estricto sentido, las enfermedades son presumiblemente asintomáticas en su etapa de latencia, como el cáncer, las enfermedades cardiovasculares o la derivada de la infección por virus de VIH, se puede afirmar que como consecuencia de esta situación, actualmente existe una

pérdida de oportunidad de realizar un diagnóstico temprano al esperar una manifestación clínica (National Research Council et al. 2012).

Aunado a lo anterior, en los modelos tradicionales de investigación mediante ensayos clínicos, los médicos tienen pocas conexiones con laboratorios de investigación e investigadores. Por lo que estos últimos, recurren a redes informales para contactar con profesionales de la salud que atienden pacientes con la enfermedad de interés a estudiar. En este mismo ámbito y como procedimiento de manera general, se enrola al paciente y se obtiene una muestra de él o se recolecta su información clínica para posteriormente pasar estos datos a un entorno de investigación, en donde se traducirá en información descriptiva o resultados que no tendrán relevancia para la mayoría de las poblaciones de pacientes debido a su heterogeneidad (National Research Council et al. 2012).

De acuerdo con la revisión realizada por la *National Academic of Sciences* en Estados Unidos se han encontrado diversas razones por las cuales los diseños tradicionales de investigación no concuerdan con las necesidades actuales, por ejemplo:

- Requiere grandes tamaños de muestra, por lo que muchos de los estudios no cuentan con el suficiente poder estadístico.
- Implican costos elevados al requerir estructuras especializadas durante el desarrollo de la investigación, por ejemplo, el enrolamiento de pacientes, el monitoreo constante de los sitios de investigación, el pago de personal administrativo, los gastos en procedimientos clínicos, así como la atención terapéutica.

- ❑ Promueven sistemas de investigación cerrados, es decir, con poco o nulo enfoque en compartir las muestras o datos de los pacientes que pudiesen ser de utilidad en otras investigaciones.
- ❑ Los investigadores y médicos desempeñan sus actividades en ambientes separados desde etapas muy tempranas de su formación, limitando la conjunción de sus conocimientos.
- ❑ No se realiza seguimiento a los pacientes una vez encontrada la correlación causal, para mejorar el diagnóstico y tratamiento a largo plazo.
- ❑ No se proporciona una retroalimentación eficiente de los resultados clínicos relevantes para integrarse con el cuidado clínico del paciente.

Teniendo en cuenta estos antecedentes, surge la necesidad de integrar toda la información biomédica y clínica disponible en nuevo modelo de medicina: la **medicina de precisión**. Esta tiene como objetivo utilizar los indicadores basados en la biología de la enfermedad como las mutaciones genómicas, biomarcadores y otros metabolitos, como indicadores potenciales que puedan medirse con un ensayo bioquímico estandarizado o de secuenciación genómica y que correlacionados con la historia clínica convencional, histológica y los hallazgos de laboratorio, puedan brindarnos la oportunidad de hacer una descripción y clasificación precisa de la enfermedad, así como proporcionar un tratamiento específico que tenga una mayor probabilidad de éxito. (National Research Council et al. 2012).

Para llevar esto todo esto a cabo, son necesarias las constantes colaboraciones internacionales, donde los investigadores deben realizar una rotación del liderazgo, autoría y recursos en la investigación, esto con la finalidad de asegurar que los beneficios de la medicina de precisión contribuyan al desarrollo de todos los países,

así como el establecimiento de infraestructuras e instalaciones científicas internacionales y no únicamente limitarse a solicitar acceso a muestras o datos, para así, incentivar la aceleración de la investigación en cada uno de las naciones participantes (Denny and Collins 2021).

2 Implementación de la medicina de precisión

Como concepto, la “medicina de precisión” (MP) se dio a conocer a partir de una publicación en 2011 del *US National Research Council* en donde se plantea la necesidad de establecer una nueva taxonomía de enfermedades, que establezca su clasificación teniendo como base el conocimiento molecular de la enfermedad (Olivier et al. 2019; Ashley 2016).

Barack Obama, en 2015 como presidente de los Estados Unidos de América (EUA), expuso durante un discurso, la visión acerca de una iniciativa nacional referente a la medicina de precisión en los Estados Unidos, otorgándole un gran foco de atención al tema. Por lo cual, la MP ha cobrado gran relevancia y está siendo promovida desde hace algunos años por diversas naciones. Por ejemplo, en Reino Unido se lleva a cabo un proyecto denominado *100,00 Genomes Project* y a la par de este, en Estados Unidos se desarrolla otro programa llamado *All of Us*. Ambos proyectos contemplan la colección de información disponible de salud de cada paciente, la construcción de la infraestructura para su procesamiento y el establecimiento de acuerdos organizacionales que permitan un intercambio de los datos generados (Minari, Brothers, and Morrison 2018).

En términos generales la medicina de precisión, busca adaptar el tratamiento médico a las características de cada paciente mediante la clasificación de los individuos en subpoblaciones, ya sea por susceptibilidad a una enfermedad, evento clínico o

resultado, y/o respuesta específica a un tratamiento, logrando que las intervenciones preventivas o terapéuticas puedan estar concentradas en quienes se benefician de ellas, permitiendo ahorrar gastos y evitar exposición a eventos adversos a quienes no lo hacen (National Research Council et al. 2012)

Cabe destacar, que el objetivo no es necesariamente generar un tratamiento único para cada paciente o individuo, sino brindar tratamientos adecuados para los subgrupos de pacientes que comparten características biológicas moleculares de la enfermedad (Seymour et al. 2017)

La propuesta de medicina de precisión va más allá de la medicina reactiva, la cual supone solamente actuar hasta que la enfermedad es manifestada mediante signos y síntomas. La MP se establece bajo 4 principales premisas que son listadas a continuación:

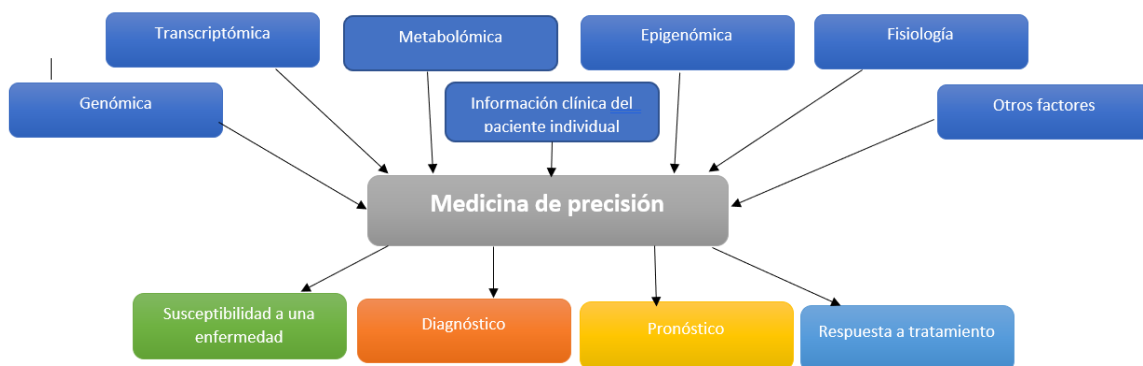
1. **Predictiva**, es decir, debe anticipar la aparición de la enfermedad evaluando factores de riesgo, estilo de vida y determinantes sociales.
2. **Preventiva**, debe proponer acciones que permitan retrasar la evolución de una patología antes de las primeras manifestaciones y una prevención secundaria una vez que la enfermedad se haya establecido.
3. **Personalizada**, debe recomendar la mejor estrategia terapéutica para la condición analizada, en función de la información genética, molecular y los factores particulares de cada individuo.
4. **Participativa**, requiere el involucramiento de la investigación biomédica, instituciones académicas, profesionales de la salud y el paciente (Crisci et al. 2020).

Es de remarcar que, para la medicina de precisión, el valor de hacer un diagnóstico preciso es la suma tanto del valor ganado en la precisión diagnóstica, pronóstica, el tratamiento personalizado y la información predictiva o de respuesta (Seymour et al. 2017).

La figura 1 esquematiza algunos de los elementos científicos de la MP y proporcionan un panorama general del modelo y los resultados esperados de esta práctica.

Figura 1.

Elementos científicos del modelo de medicina de precisión.



Esquema que presenta los elementos científicos y disciplinas necesarias para la construcción del modelo de medicina de precisión (recuadros azul oscuro, parte superior) así como los resultados de esta práctica (recuadros de colores en la parte inferior).

Fuente: *Building the foundation for genomics in precision medicine* (p. 1439-1451), por Love-Koh et al., 2018. PharmacoEconomics.

Ya se ha mencionado el objetivo de la MP, por lo que ahora se mencionan los requerimientos generales para implementar este modelo, los cuales podrían resumirse mediante los siguientes puntos:

- Ø Desarrollo de una **nueva taxonomía de las enfermedades** que esté basada en la biología intrínseca, es decir, en las causas moleculares de la enfermedad, adicional a los signos y síntomas.

∅ Adición de una **descripción** que esté directamente relacionada con la comprensión de los mecanismos de la enfermedad, patogénesis y tratamientos.

∅ **Incorporación de información clínica** referente a pacientes individuales, esta información puede ser generada durante la consulta ordinaria y reportada mediante el EHR (*Electronic Health Record*, por sus siglas en inglés).

∅ Integración de múltiples parámetros en una sola **red de información común** que esté disponible para los investigadores en todo momento y que permita la continua incorporación de nueva información acerca de las enfermedades.

De la misma manera, se requiere de una estrecha colaboración entre los investigadores y el personal de salud para la implementación de este nuevo modelo, visualizando la MP en coexistencia de la atención clínica. Se propone como un mismo sistema, en el que se coordinan todas las actividades necesarias, algunas de las cuales se mencionan a continuación: a) búsqueda de casos clínicos, b) adición constante de la nueva información proveniente de las tecnologías ómicas en una red común de información, c) utilización de los datos disponibles en los EHR, y d) enrolamiento de pacientes en ensayos clínicos. Estas acciones son las principales, pero no son limitantes de las restantes (Seymour et al. 2017)

Para llevar a cabo estas actividades es fundamental el desarrollo de un ecosistema que involucre al personal de salud, hospitales, instituciones públicas de investigación, investigadores, laboratorios, empresas de investigación, bioinformáticos, sistemas de información clínica y autoridades regulatorias en materia sanitaria. En este ecosistema, es también importante la construcción de biobancos, que vinculen las muestras con los datos de salud de los pacientes para apoyar los hallazgos relevantes. Así mismo, es necesario desarrollar bases de datos de búsqueda que

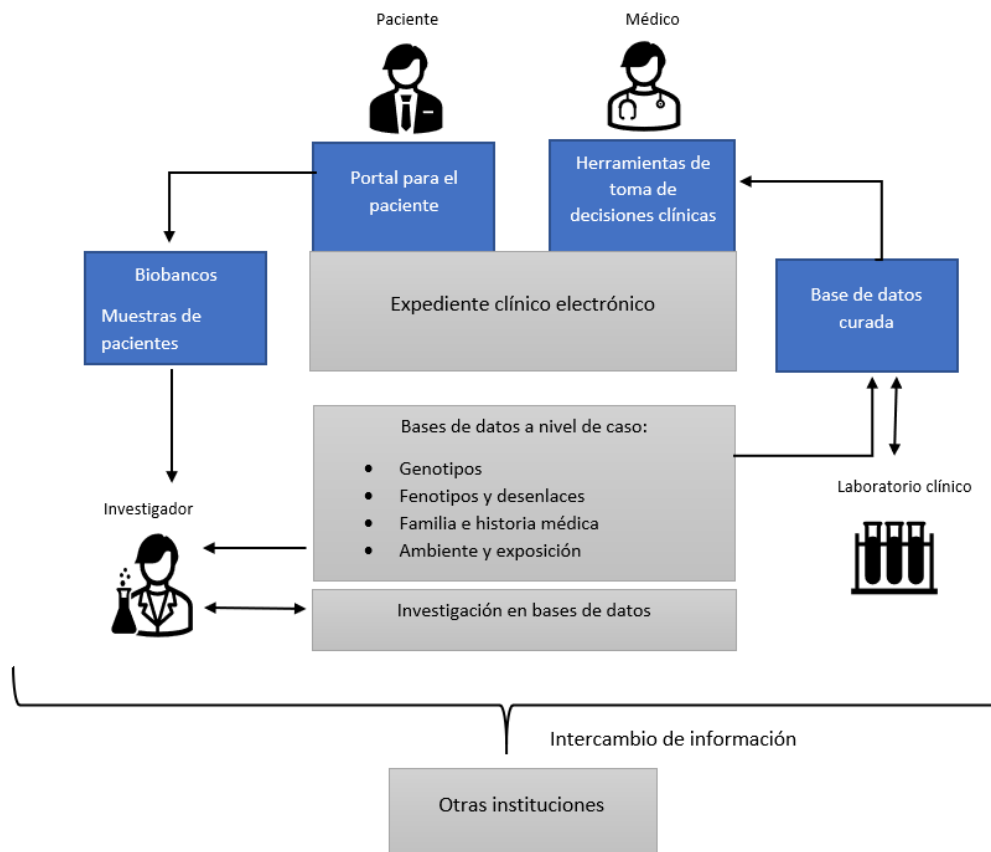
almacenen la información, cálculos y resultados que suministrarán la evidencia para una interpretación clínica.

Posteriormente, entran en juego las redes de intercambio de conocimientos que permiten compartir esta información y perfeccionar las interpretaciones. Los laboratorios clínicos y sus sistemas de información facilitan la consolidación de interpretaciones clínicas en forma de informes y alertas. Los EHR y los sistemas asociados, ayudan a los médicos en la aplicación práctica de los resultados, a medida que evoluciona la condición del paciente y el conocimiento de las variantes implicadas en la predisposición, desarrollo y tratamiento de la enfermedad.

Con base en lo descrito anteriormente, se esquematiza en la **figura 2** el ecosistema de la medicina de precisión. En el esquema se muestra cómo se conectan pacientes, médicos, personal de atención clínica, investigadores y laboratorios como principales actores. Los pacientes y médicos tendrán acceso a la información de los EHR mediante portales o herramientas de decisión clínica. Los biobancos reciben datos de casos o muestras provenientes de flujos de investigación, beneficiando a los investigadores al acceder a todas estas fuentes de información y contribuyendo a su vez, al enriquecimiento de estas fuentes. Como paso final, los laboratorios clínicos aprovecharán estos datos para informar a la comunidad clínica acerca de los hallazgos y cuál es su impacto en la salud.

Figura 2.

El ecosistema de la medicina de precisión.



Fuente: *Building the foundation for genomics in precision medicine* (p. 337) por Aronson and Rhem (2015) Nature.

Aunque la MP ya se practica en diferentes sitios de reconocimiento internacional como el Instituto de Medicina Genómica Geisinger, Scripps Health, la Clínica Cleveland, el Centro Médico de la Universidad de Vanderbilt y el Colegio Médico de Wisconsin, por nombrar solo algunos lugares, aún se considera un proyecto a largo plazo, pues este ha empezado a construirse con una generación y podría ser completado por otra, en donde la planificación y construcción puede cambiar conforme nuevas tecnologías sean involucradas (Alzu'bi, Zhou, and Watzlaf 2014).

Este documento se centra en revisar y describir el estado actual de los diferentes aspectos que se necesitan para llegar a la medicina de precisión, considerando su impacto en el desarrollo de sistemas de salud en el futuro, y a su potencial para

mejorar la salud individual y colectiva de la población, enfatizando el contexto de los retos éticos, logísticos y tecnológicos que la acompañan.

3 Desarrollo de tecnologías de alto rendimiento con aplicación médica

La mayor parte de las publicaciones se refieren a la MP con un enfoque genómico, lo cual se debe a que la tecnología genómica fue la primera tecnología en tener información disponible para la aplicación del modelo de medicina de precisión. En estudios recientes, la secuenciación de todo el genoma y todo el exoma han proporcionado un gran conjunto de variantes del DNA que son asociadas con distintas enfermedades. Pero, a pesar del rápido crecimiento del conjunto de datos, no se ha podido establecer la relevancia clínica exacta de todas las variantes descubiertas. A pesar de ello, si se ha podido establecer algunas escalas de riesgo poligénico en un rango de enfermedades comunes y con una importante aplicación clínica, sin necesidad de dilucidar la secuencia causal individual (Aronson and Rehm 2015).

De la misma manera, en las dos últimas décadas se han desarrollado equipos y técnicas que permiten el análisis simultáneo de muchas moléculas, es por ello que la cantidad de información molecular, genómica, proteómica y metabolómica de los seres humanos ha incrementado considerablemente. Algunos estiman que desde el 2002 el volumen de datos de secuenciación genética se ha multiplicado por diez cada año. Esto genera la necesidad de aprovechar estos datos mediante un nuevo modelo de medicina en beneficio de los pacientes (National Research Council et al. 2012; Seymour et al. 2017)

Las tecnologías de secuenciación de alto rendimiento nos han ayudado a descifrar el código genético de miles de personas. Por su parte, las tecnologías de imagen y microscópicas nos ayudado a visualizar y monitorear los procesos en células y tejidos

y otros avances t en técnicas como la NMR (*Nuclear Magnetic Resonance*, por sus siglas en inglés), la espectrometría de masas y los ensayos basados en metodologías de microfluidos, han contribuido a conocer la composición molecular y las funciones de los sistemas biológicos. Toda esta información es la base para comprender y caracterizar los mecanismos moleculares que están involucrados en el desarrollo y progreso de las enfermedades (Olivier et al. 2019)

El ágil desarrollo de estas tecnologías ha contribuido a disminuir los costos de manera considerable. Por otra parte, los avances informáticos nos han permitido generar información clínica detallada acerca de un extenso número de pacientes. Así mismo, el cambio de la actitud pública hacia el acceso de los datos moleculares y la privacidad de los datos de salud ha permitido realizar investigación básica y aplicada a partir de toda la información disponible (Ashley 2016; National Research Council et al. 2012).

4 Investigación básica

Además de la información clínica y epidemiológica relevante, la medicina de precisión se basa en la información proporcionada por varios campos ómicos que han surgido en las últimas décadas: a) la **genómica**, que estudia el genoma o sus subconjuntos como son el exoma entre otros, b) la **transcriptómica**, que se ocupa del conjunto completo de transcripciones de una célula, tejido u organismo, c) la **proteómica**, que comprende el estudio del conjunto completo o un subconjuntos de proteínas presentes en un tipo de célula o tejido, d) la **epigenómica**, que investiga el conjunto completo de modificaciones covalentes del ADN que no alteran la secuencia del ADN en sí, pero que producen cambios en la actividad genómica y su procesos de regulación, e) la **metabolómica**, que analiza el conjunto completo de metabolitos de bajo peso molecular (p. ej., aminoácidos, ácidos orgánicos, lípidos y azúcares), f) la

farmacogenómica, la cual se refiere al estudio de la contribución de las diferencias en los genes de un individuo a la variación en las respuestas a los medicamentos entre la población y g) el campo que estudia el **exposoma**, que comprende moléculas y eventos a los que está expuesta una persona (por ejemplo, medicamentos, dieta y otros factores ambientales)

4.1 Genómica

El genoma consiste en el conjunto de genes que especifican todos los caracteres que pueden ser expresados por un organismo y la genómica por lo tanto es la disciplina que se encarga del estudio de los genes y su expresión. Conocer el genoma humano contribuye a comprender la complejidad de las enfermedades, por lo cual el proyecto del genoma humano HGP (*Human Genome Project*, por sus siglas en inglés) fue un hito en la historia de las ciencias biológicas. Este proyecto nació en 1990 con el objetivo de identificar los genes que configuran nuestro patrimonio genético (Nature 2021).

Durante los primeros años de este proyecto se desarrollaron estrategias que permitieron identificar genes y su posición relativa en el genoma por medio de mapas genéticos y análisis de la secuencia de nucleótidos en el ADN con mapas físicos. En 2004, en una primera fase, el proyecto cumplió el objetivo de secuenciar el genoma humano con una precisión de error de una base por cada 100,000 bases, así como se determinó que dicho genoma es 99.9 % idéntico entre individuos y la diferencia restante se compone de variaciones denominadas polimorfismos: SNP (*Single Nucleotide polymorphism*, por sus siglas en inglés) que consisten en cambios puntuales de nucleótidos. La versión actual del genoma humano de referencia, GRCh38.p13, comprende 3.270 millones de nucleótidos (Nature 2021).

En una segunda fase e impulsado por el proyecto HapMap (Haplotype map, por sus siglas en inglés) se hizo posible la caracterización de la variación común en todo el genoma, cuando cientos de miles de marcadores se pudieron analizar simultáneamente en plataformas de *microarrays* (microarreglos). Más adelante, las técnicas de secuenciación de nueva generación permitieron el estudio del genoma completo, donde un secuenciador lee de forma ordenada los nucleótidos de los diferentes fragmentos del ADN del genoma. Durante este proceso, el ADN se fracciona en pequeños segmentos para posteriormente realizar la lectura de cada segmento. Estas lecturas pueden ensamblarse para obtener el genoma completo o bien, pueden ser comparados con un genoma de referencia, dándonos como ventaja la secuenciación de una gran cantidad de ADN en un periodo corto y la detección de variantes puntuales, así como inserciones y deleciones (cortas o largas). Esto derivó, en cambios en los objetivos del análisis del genoma humano, los cuales están descritos en la Tabla 1 (McGrath and Gherzi 2016).

Tabla 1.

Tabla comparativa de objetivos de estudio en la genómica, antes y después del HGP.

Objetivos de investigación en variaciones genéticas y estudios de asociación de enfermedades antes y después del Proyecto del Genoma Humano	
Antes del Proyecto del genoma humano	Después del Proyecto del genoma humano
Mutaciones en uno o pocos genes	Variaciones genéticas en varios genes
Genes codificantes para proteínas	Genes codificantes para proteínas y RNA
Regiones en un gen codificantes para proteínas	Regiones intragénicas (intrones); elementos reguladores, regiones intergénicas
Principales mutaciones	SNP; inserciones y deleciones cortas, variaciones en el número de copias, otras variaciones estructurales.
Genes individuales	Interacciones gen-gen, interacciones ambiente-gen
Estudios de un gen o unos pocos genes por estudio.	Decenas de miles de genes de genes o el estudio del genoma completo en un solo estudio.

FUENTE: (McGrath and Gherzi 2016)

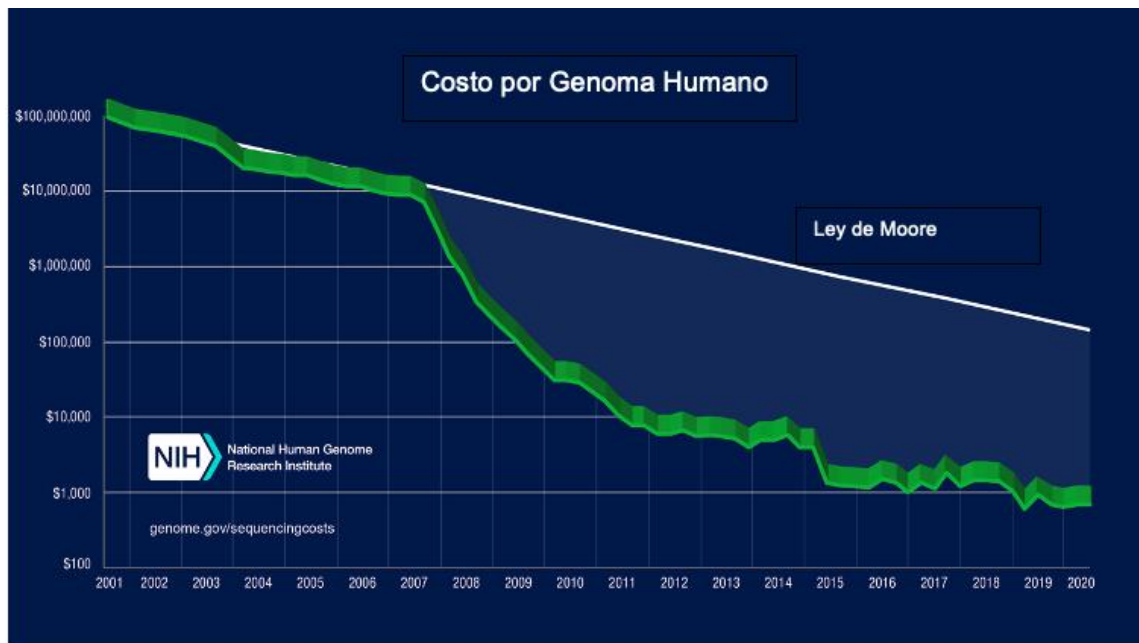
También debemos considerar el cambio en los costos de secuenciar el genoma, pues este depende de la cobertura de la secuenciación. La cobertura de la secuenciación es el número promedio de veces que se lee cada una de las bases a lo largo del genoma y determina la capacidad de encontrar variantes y la certidumbre con la que determinamos los genotipos correctos. Donde una secuenciación de alta cobertura nos otorgará mayor certeza, pero a su vez incrementa el costo y en el caso contrario cuando reducimos la cobertura, reducimos el precio, pero con ello la certeza de la información. Es por esto que el objetivo es encontrar aquellas estrategias de secuenciación que permitan maximizar el número de individuos secuenciados con una cobertura y costo adecuado al proyecto de investigación en curso. Otra ventaja es que, aunque a medida que se disminuye la cobertura se reduce el índice de detección de variantes a nivel individual, pero a nivel poblacional el índice de detección de variantes se incrementa al aumentar el número de individuos estudiados (Samperi Tena 2020)

El proyecto del genoma humano es un ejemplo de cómo la colaboración internacional a gran escala puede generar recursos de invaluable valor científico como el establecimiento de un genoma de referencia, el cual tuvo un costo de 2.7 billones de dólares aproximadamente. Actualmente, una secuenciación puede realizarse por menos de 500 dólares, representando un paso en la asequibilidad y estableciendo una disminución en tiempo y costo (Figura 3), que han permitido aumentar el análisis de la variación genética entre individuos, dando una ventaja al desarrollo de la MP, con lo cual se da paso a la genómica personal o acceso a la secuenciación de manera individual, pues el genoma de referencia es un compuesto de varios genomas donantes anónimos y no representa a ningún individuo. Sin embargo, cuando se extiende la posibilidad de la secuenciación del genoma a individuos, se permite a los

participantes no sólo conocer información de sus ancestros, sino también la que se encuentre asociada al desarrollo de enfermedades para ellos mismos o sus familiares (Samperi 2020).

Figura 3.

Costo de la secuenciación del genoma humano a través del tiempo hasta 2020.

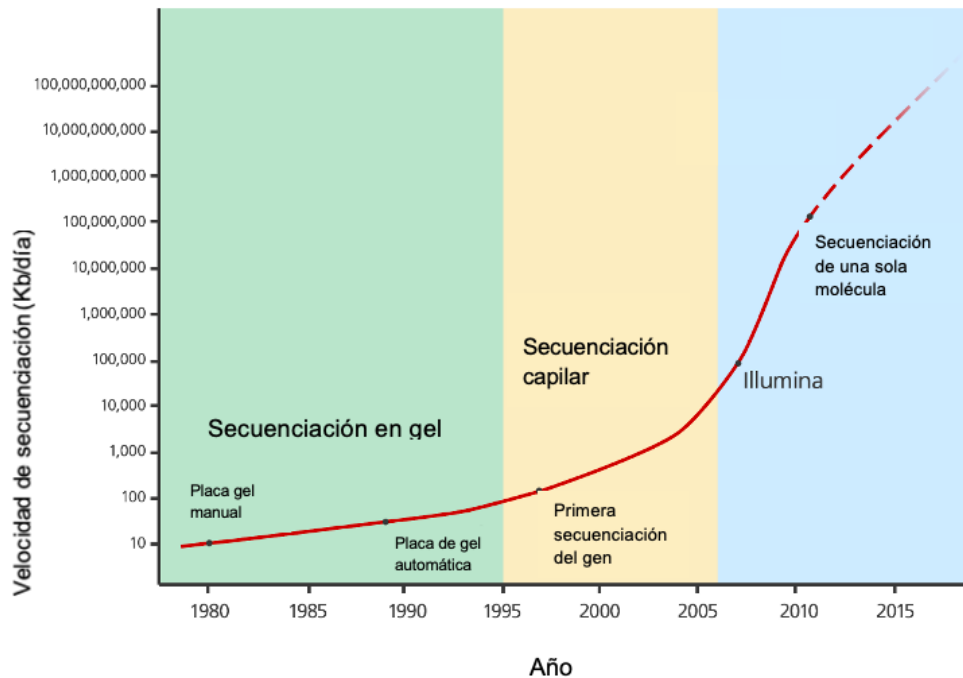


El costo de la secuenciación completa del genoma está cayendo más rápido que la Ley de Moore. Fuente: <https://www.genome.gov/about-genomics/fact-sheets/Sequencing-Human-Genome-cost> (Fecha de consulta: Oct-2021)

Es por ello, que el surgimiento y la evolución de las plataformas de secuenciación del genoma completo (WGS, por sus siglas en inglés) condujo a la identificación rápida y una comprensión más completa de las fuentes de variación genética: reordenamientos cromosómicos, variación del número de copias de genes, cambios epigenéticos y polimorfismos de un solo nucleótido (Figura 4) (Sandhu, Qureshi, and Emili 2018).

Figura 4.

Evolución sobre las técnicas de la secuenciación.



En las dos primeras regiones del gráfico se muestran las tecnologías basadas en los métodos de Sanger. Sobre la primera década del 2000 apareció la secuenciación masiva. Es importante observar cómo ha ido aumentando el rendimiento (línea roja). Fuente: Estrategias de identificación de variantes asociadas a caracteres productivos mediante chips o secuenciación a baja cobertura por (Samperi 2020)

Antes del HGP, secuenciar un genoma era difícil y consumía mucho tiempo, por lo tanto, los científicos sólo podían centrarse en analizar un gen o unos pocos genes en sus proyectos de investigación. En consecuencia, los investigadores asociaron que la mayoría de las enfermedades genéticas eran causadas por mutaciones en un gen o algunos genes. Sin embargo, la investigación posterior al HGP ha indicado que esta comprensión no era correcta (Hood and Rowen 2013).

Las asociaciones entre la enfermedad y la información genómica se han extendido de un solo gen a múltiples genes en uno o múltiples cromosomas. Esto incluye sectores como: regiones codificantes de proteínas de un gen hasta las regiones intragénicas (intrones, los cuales son no codificantes para proteínas), elementos reguladores hasta

las regiones intergénicas, c) otros tipos de ARN (por ejemplo, ARNt, microARN), polimorfismos de un solo nucleótido (SNP), inserciones y deleciones, variaciones en el número de copias (CNV, *copy number variations* por sus siglas en inglés) y otras variaciones estructurales, interacciones gen-gen, interacciones gen-ambiente; y de un solo o unos pocos genes por estudio hasta estudios de asociación de todo el genoma (WGAS, *Whole Genome Association Study*, por sus siglas en inglés), los cuales desde principios de los 2000 han sido una de las principales herramientas para el descubrimiento de los genes y los mecanismos genómicos de las enfermedades (Glicksberg, Johnson, and Dudley 2018).

Por otro lado, el descubrimiento de las SNP permitió realizar la secuenciación de baja cobertura en grandes poblaciones, y estas a su vez dieron lugar a la deducción de algunas relaciones de asociación entre rasgos fenotípicos y la presencia de ciertas SNPs, en su mayoría en regiones no codificantes del genoma. Dichas variantes han cobrado relevancia al correlacionarse con la incidencia de enfermedades jugando un rol importante en la identificación de genes de susceptibilidad a enfermedades (diagnóstico predictivo) y como indicadores de respuesta al tratamiento o fármacos (Farmacogenómica) (“Genotipado En La Salud Humana : Informe de Vigilancia Tecnológica” n.d.) siendo el objetivo principal el caracterizar tales mutaciones para identificar dianas moleculares para la terapia.

A pesar de este tipo de enfoque dirigido, todavía existen algunos pacientes que no responden a la terapia o que presentan solo una respuesta parcial a pesar de que la mutación estudiada este presente. Esto probablemente refleja la presencia de mutaciones adicionales no detectadas en la red bioquímica y refleja las limitaciones de dirigirse a una sola molécula. Es aquí, donde los análisis de la *big data* genómica

pueden tener el potencial de generar hipótesis que no son evidentes a partir de datos de secuencias individuales (Sandhu, Qureshi, and Emili 2018).

Respecto al desarrollo de la genómica y de acuerdo con Aronson and Rehm (2015) en la actualidad y en el ámbito clínico, los médicos tienen acceso a una mayor variedad de pruebas que les permiten determinar las variantes genómicas que existen en sus pacientes. Estas pruebas incluyen:

§ **Genotipado de variantes**, que se sabe no están asociadas con efectos clínicos importantes

§ Secuenciación **en panel**, que analiza genes que están relacionados con una indicación específica para detectar variantes nuevas o conocidas.

§ Secuenciación **del exoma**, todos los genes codificantes conocidos.

§ **Análisis de la secuenciación del genoma completo.**

El entendimiento de la relación entre las características moleculares y fenotípicas nos ayudará en el desarrollo de estrategias para comprender las implicaciones individuales de cada paciente.

Por ejemplo, el descubrimiento de ciertas mutaciones en los genes BRCA 1 y 2 permite identificar el riesgo de desarrollar cáncer de mama y manejar este riesgo mediante cirugías profilácticas o monitoreo constante de los pacientes (National Research Council *et al.*, 2012).

Actualmente el campo de mayor aplicación de estas técnicas es la oncología, donde los estudios genómicos estaban enfocados en la identificación de mutaciones heredadas que incrementaban significativamente el riesgo de cáncer, tomando como

base estudios de familias en donde varios miembros presentaron un tipo particular de cáncer, llevando los estudios hacia la identificación de varios genes y mutaciones, que cuando son heredadas aumenta el riesgo de padecer tipos específicos de cáncer.

Por ejemplo, el Síndrome de Lych, el cual es un tipo de cáncer colorrectal sin poliposis, es frecuentemente ocasionado por mutaciones en genes implicados en la reparación de errores de apareamiento del ADN (MLH1, MSH2, MSH6, PMS1 y PMS2), otro ejemplo son las mutaciones en el gen supresor de tumores TP53 que puede tener como consecuencia el desarrollo de distintos tipos de cáncer como sarcoma o leucemia. Llegando a la conclusión que los individuos que presentan mutaciones en los genes mencionados tienen un riesgo significativamente más alto de desarrollar la enfermedad comparado con los que no lo portaban. Esto fue resultado de los estudios tempranos de la asociación del genoma completo, pero el nuevo enfoque se basa en evaluaciones mediante puntuaciones de riesgo poligénico que ya incluye el número de variantes y mutaciones, que ahora han sido asociadas con el incremento de riesgo de padecer cáncer (Olivier *et al.*, 2019).

Recientemente la secuenciación del ADN se han enfocado en identificar mutaciones específicas del cáncer mediante el análisis de diversas variantes provenientes de células cancerosas como SNP, inserciones o deleciones, variaciones en el número de copias y rearrreglos estructurales, pero el número total de estas mutaciones varía ampliamente entre tipos de tumores y entre tumores iguales pero provenientes de diversos individuos, por lo cual es compleja la identificación de las variantes claves para el desarrollo de la enfermedad (Olivier *et al.*, 2019)

En la actualidad, un número significativo de estas variantes se determinan de forma rutinaria como diagnósticos complementarios para la caracterización y estadificación

de tumores y la toma de decisiones sobre el tratamiento. A pesar del progreso en las pruebas clínicas para detectar formas de cáncer hereditarias, se busca que los crecientes conjuntos de datos disponibles revelen el perfil subyacente de las mutaciones somáticas acumuladas en células tumorales para poder distinguir las *driver mutations* (responsables de la formación y crecimiento tumoral) de las *passenger mutations* (fenotípicamente neutrales) (Olivier *et al.*, 2019).

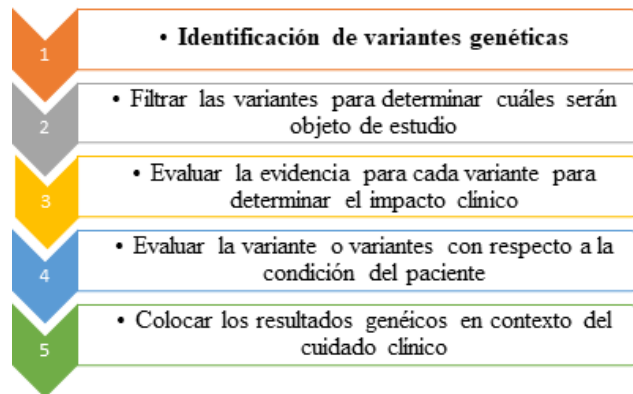
Otras utilidades que se han encontrado entre la correlación de las características genómicas de ciertas subpoblaciones y de las cuáles se benefician los pacientes una vez correctamente estatificados son: la determinación de la variabilidad en la respuesta a fármacos, la prevención de efectos adversos específicos y la dosificación adecuada según el genotipo, esto mediante la farmacogenómica, tema que será abordado más adelante en este mismo trabajo.

Es importante mencionar que las pruebas genómicas constan de dos componentes: Uno técnico que está basado en identificar qué variantes están presentes y un componente interpretativo en el que se evalúan las implicaciones de las variantes identificadas.

Dado que la interpretación es crucial se han propuesto metodologías que ayuden a esta tarea., como se muestra en la figura 5. Como primer paso y posterior a la identificación de las variantes, se debe filtrar cual de ellas se debe investigar más a detalle. En segundo lugar, se deberá evaluar la evidencia para cada variante y determinar cuál es su impacto clínico. Como tercer paso se evalúa la variante o variantes con respecto a la condición del paciente. Como último se colocan los resultados en contexto de toda la información conocida acerca del paciente para determinar cual deberá ser la atención clínica más adecuada.

Figura 5.

Proceso general de la interpretación de las variantes genómicas



Etapas del proceso de interpretación genética.

Una vez identificadas las variantes genéticas (paso 1), se filtran para seleccionar las de interés (paso 2). A continuación, se evalúa la evidencia de cada variante para determinar el impacto clínico de la variante (paso 3). A continuación, se interpretan una o más variantes con respecto a la afección específica por la que se está investigando al paciente (paso 4). Por último, la evaluación genética general se coloca en el contexto clínico y personal del paciente para informar el proceso de toma de decisiones de atención clínica (paso 5).

Fuente: *Building the foundation for genomics in precision medicine* (p. 337), por Aronson and Rehm (2015).

Se debe considerar, que la interpretación tiene como base; el juicio profesional, por lo cual está sujeto a los errores humanos y las diferencias de opinión.

A pesar de que los estudios del genoma han revelado proporciones de riesgo de enfermedad bastante modestas, aunque estadísticamente significativas, se necesita investigar las interacciones entre factores genómicos y no genómicos para identificar las circunstancias de riesgo de la enfermedad que no se evalúan en los estudios convencionales. Es decir, los parámetros moleculares no deben ser limitados a como actualmente son entendidos, se deben considerar datos ambientales, de comportamiento y factores socioeconómicos (National Research Council *et al.*, 2012).

Por lo que, al ser una necesidad la identificación de los individuos con alta probabilidad de desarrollar una enfermedad y tomando en consideración que la

mayoría de las enfermedades comunes tienen un componente genético, es que, la estatificación de los individuos en función de la variación del ADN se ha convertido en un enfoque fundamental.

Para lograr esta detección de individuos en riesgo se han desarrollado diversas propuestas centradas principalmente en la búsqueda de portadores de mutaciones monogénicas, a pesar que se ha determinado que las enfermedades más comunes conllevan un riesgo de naturaleza poligénica, , la cual está definida como la evaluación del riesgo de una afección específica teniendo en cuenta la influencia colectiva de muchas variantes genéticas. Entre ellas, variantes asociadas a genes con funciones conocidas y variantes que no se sabe que están asociadas a genes relevantes para la afección.

Khera *et al.*, (2018) en un artículo donde analizan que una puntuación poligénica para enfermedades comunes pueden identificar un riesgo equivalente al de mutaciones monogénicas. En él, evalúan y establecen que de la población del estudio: 8.0 % tiene un riesgo superior al triple de padecer una enfermedad arterial coronaria, 6.1% fibrilación auricular, 3.5 % diabetes tipo 2, 3.2 % una enfermedad inflamatoria intestinal, y 1.5 % cáncer de mama. Por lo que proponen incluir la predicción del riesgo poligénico en la atención clínica, concluyendo que los puntajes de riesgo poligénico pueden lograr un pronóstico similar al de las variantes raras de la enfermedad mendeliana, las cuales por sí solas representan una probabilidad muy alta de desarrollar la enfermedad.

El uso de puntajes de riesgo poligénico puede permitir a los proveedores de servicios de salud estatificar el riesgo de individuos que de otro modo serían ignorado, y así identificar nuevas poblaciones para tratamiento y mejorando un filtro para la

realización de pruebas a una población aparentemente sana e identificar enfermedades que aún no muestran manifestaciones clínicas.

Si todas las personas tuvieran su secuencia genómica anexa a su EHR, se tendría al alcance mejores alternativas terapéuticas al contrastar la información disponible para una adecuada prescripción en cada paciente, destacándose el uso de terapias guiadas por la genómica como el estándar de cuidado en varios tipos de cáncer usando las fármaco variantes de líneas germinales siendo actualmente adoptadas por aún pocos centros en Estados Unidos debido a la falta de EHR habilitados para el registro de los datos genómicos.

4.2 Transcriptómica

La caracterización de genes ha sido un objetivo primordial para los investigadores en genética y genómica. Sin embargo, la medición directa del producto génico, el ARN transcrito, proporciona una lectura funcional del genoma que es esencial para nuestra comprensión del desarrollo de la enfermedad. El transcriptoma es el conjunto total de RNA transcrito, el cuál esta dividido en codificante (1–4 %—ARN mensajero) y no codificantes (>95 %—ribosómico, de transferencia, nuclear pequeño, de interferencia pequeña, micro y no codificante largo) (Manzoni et al. 2018)

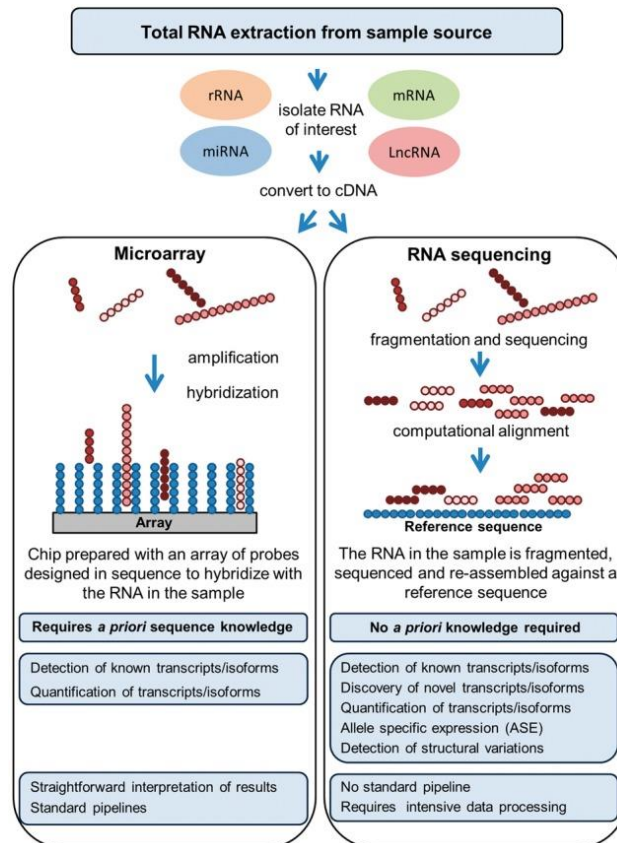
A principios de la década de 2000, se utilizaron enfoques basados tanto en la hibridación como en la secuenciación, para la cuantificación del ARN de todo el transcriptoma. Las tecnologías de hibridación conocidas como microarreglos, tienen algunas limitaciones, pues las matrices de genes se construyen sobre diseño. Esto es debido a que solo se incorporan genes o exones conocidos, por lo cual no es posible hacer nuevos descubrimientos. Tienen además una sensibilidad limitada para detectar genes de muy baja expresión, así como diferenciar entre genes con

homología de secuencia (especificidad). Por el contrario, los enfoques basados en secuencias, como el análisis en serie de la expresión génica (SAGE, *Serial Analysis of gene expression*, por sus siglas en inglés) o la secuenciación masiva de firmas paralelas (MPSS, *Massively parallel signature sequencing*, por sus siglas en inglés), se fundamentan en la secuenciación de Sanger y presentaban una capacidad limitada para detectar todas las isoformas de la transcripción (Zhong Wang, Gerstein, and Snyder 2009).

La aparición de las tecnologías de secuenciación de ADN de próxima generación, permitió el desarrollo de metodologías de secuenciación de alto rendimiento de transcriptomas completos, conocida como secuenciación de ARN (RNA-seq, RNA Sequencing por sus siglas en inglés). Un experimento típico de RNA-seq consiste en aislar el ARN mensajero (mRNA) con una cola poli (A), transcribir inversamente el RNA en ADN complementario (cDNA), secuenciar este cDNA usando un instrumento de secuenciación de próxima generación y mapeando las lecturas resultantes al genoma de referencia (Figura 6). RNA-seq permite identificar qué partes del genoma se transcriben y cuantificar el ARN con una mayor sensibilidad, al detectar transcritos de baja expresión e identificar nuevos genes, exones e isoformas (Zhong Wang, Gerstein, and Snyder 2009).

Figura 6.

Resumen del análisis de datos de secuenciación de ARN mediante microarreglos y RNA sequencing.



Fuente: **Genome, transcriptome and proteome: the rise of omics data and their integration in biomedical sciences** por Claudia Manzoni et al (2018) Briefings in Bioinformatics.

El análisis de ARNm proporciona información directa sobre características de expresión génica específicas de células y tejidos, como:

- Presencia/ausencia y cuantificación de una transcripción
- Evaluación de empalmes alternativos/diferenciales para evaluar o predecir isoformas de proteínas
- Evaluación cuantitativa de la influencia del genotipo en la expresión génica mediante análisis de loci de rasgos cuantitativos de expresión (eQTL) o expresión específica de alelo (ASE).

Esta información es fundamental para una mejor comprensión de la dinámica del metabolismo celular y tisular, y para apreciar si los cambios en los perfiles transcriptómicos afectan la salud y la enfermedad y es que cómo lo hacen

La técnica RNA-seq, reveló la complejidad del genoma codificante eucariota, descubriendo muchos genes superpuestos y sitios de inicio de transcripción alternativos (Nature 2021). Hasta la fecha y como ejemplo de antes lo mencionado, se han depositado casi 800.000 conjuntos de datos de expresión génica relacionados con el cáncer en la base de datos *Gene Expression Omnibus* (GEO, por us siglas en inglés) del Centro Nacional de Biotecnología (NCBI) de los NIH (National Institute of Health).

También se pueden realizar diferentes tipos de análisis post hoc con datos de secuenciación, incluidos perfiles de expresión diferenciales específicos de células y/o tejidos que prueban si los genes tienen diferentes niveles de expresión entre tejidos y condiciones o análisis que investigan regulaciones de expresión génica como Análisis eQTL.

El mapeo de eQTL es un enfoque poderoso para investigar e integrar medidas de expresión génica (datos a nivel de ARN) con variantes genéticas (datos a nivel de ADN). El análisis de eQTL ayuda a identificar variantes genéticas que influyen en los niveles de expresión de ARNm. La relación de ambos datos la podemos realizar a través de los SNP, los cuales se encuentran en su mayoría regiones no codificantes y, por lo tanto, es probable que ejerzan su función a través de la regulación de la expresión génica y por lo tanto, se pueden utilizar para explorar y definir mejor las redes moleculares subyacentes asociadas con un rasgo particular.

Otra forma de estudiar la regulación de la expresión génica se logra a través del análisis combinado de los niveles de ARNm y microARN(Manzoni et al. 2018).

La unión específica de microARN a un ARNm objetivo inhibe la unión del ARNm al ribosoma o dirige al ARNm para su degradación, siendo entonces la presencia de cierto

microARN como un indicador predictivo de una determinada condición o rasgo y, por lo tanto, puede usarse de inmediato en el diagnóstico clínico.

El enfoque de perfilado de ARNm/microARN se ha aplicado con éxito, por ejemplo, para estudiar la expresión génica en subtipos específicos de tumores gastrointestinales o para evaluar la alteración de la expresión génica en fibroblastos con problemas de cicatrización de heridas de pacientes diabéticos (Manzoni et al. 2018).

La iniciativa *The Cancer Genome Atlas* (TCGA, por sus siglas en inglés) es un proyecto que nace con el objetivo de catalogar las mutaciones responsables del cáncer a través de las NGS (*Next generation sequencing technologies*, por sus siglas en inglés) y parte importante de ésta iniciativa es la transcriptómica pues proporciona una caracterización detallada de la expresión génica de tumores y tejidos individuales de más de 11.000 pacientes y junto con otras iniciativas como el Proyecto de la Fundación *Stand Up To Cancer-Prostate*, ha ayudado al análisis detallado de los cambios en la expresión génica asociados con la formación y el crecimiento de tumores. (Zhining Wang, Jensen, and Zenklusen 2016)

4.3 Epigenómica

El epigenoma se refiere a las modificaciones del ADN y las proteínas asociadas que ayudan en la regulación de la expresión génica. Los cambios más comunes incluyen la metilación del material genético o modificaciones de histonas, así como la expresión de miRNAs (microRNA, por sus siglas en inglés) y LncRNAs (*Long non coding RNA*, por sus siglas en inglés) y son un proceso natural influenciado por el medio ambiente o la enfermedad (Olivier et al. 2019).

El ADN se encuentra empaquetado y superenrollado junto con las histonas para formar una estructura denominada cromatina. Esta estructura nos ayuda a garantizar la estabilidad, replicación y correcta transcripción del código genético. La cromatina está ligada de manera física a múltiples proteínas que funcionan como reguladoras de la expresión o silenciamiento de los genes en las funciones celulares. Por lo que ha cobrado importancia la caracterización de cómo, cuándo y dónde se unen estas proteínas reguladoras a la cromatina (Nature 2021)

En 1980, la técnica ChIP (*Chromatin Immunoprecipitation*, por sus siglas en inglés) permitió la identificación de las regiones de ADN que están unidas a proteínas de interés. En esta técnica, se trata a las células con un agente reticulante como el formaldehído para inducir la formación de enlaces covalentes entre la cromatina y la proteínas de interés de manera que no se pierda la interacción estructural y regulatoria. Posteriormente se fragmenta la cromatina y se reconoce los fragmentos de DNA asociados con la proteína de interés mediante el uso de un anticuerpo. Una vez identificados, se realiza la ruptura de los enlaces covalentes y el DNA se purifica o aísla de proteínas incluidas las histonas. Este DNA aislado se analizaba mediante plataformas de microarrays, lo cual limitaba la resolución, es decir el grado de detalle molecular en un mapa físico de DNA (Nature 2021)

En el año 2007 y gracias a las tecnologías de secuenciación de alto rendimiento, se le adiciono un paso a la técnica inicial: la secuenciación de los fragmentos del DNA de baja cobertura, lo cual como ha sido mencionado antes nos podría llevar al descubrimiento de SNP que sean relevantes en el desarrollo de las patologías. Recientemente se han presentado alternativas innovadoras a ChIP-Seq como CUT&RUN y CUN&Tag, las cuales proporcionan un mapeo de la cromatina

optimizado utilizando profundidades de secuenciación más bajas y generando así datos de mayor calidad a un precio reducido, sin embargo, ChIP-Seq sigue siendo la metodología estándar en los estudios transcripcionales y de epigenómica, pues la adopción rápida de esta metodología por investigadores de varias áreas como la oncología ha tenido repercusión en la comprensión en la biología de la cromatina y los mecanismos reguladores de genes así como el papel de las variantes genéticas no codificantes de la enfermedad (Ashley 2016)

Un ejemplo de la importancia de las diferencias epigenéticas puede ser observado entre células tumorales y células normales, así como en células tumorales durante el progreso de la enfermedad o en la respuesta a tratamiento. En el caso de la metilación del DNA, esta se ha utilizado como una herramienta de decisión de tratamiento en algunas condiciones.

La metilación de la citosina (mC) en el contexto de los dinucleótidos CG (CpG) es la modificación de ADN más común en los mamíferos, incluidos los humanos. La mayoría de los aproximadamente 28 millones de CpG en el genoma humano están metilados, mientras que los que se encuentran en los promotores de genes generalmente ricos en CpG están en gran parte sin metilar. La metilación aberrante del ADN en las células tumorales fue una característica epigenética temprana observada en el cáncer

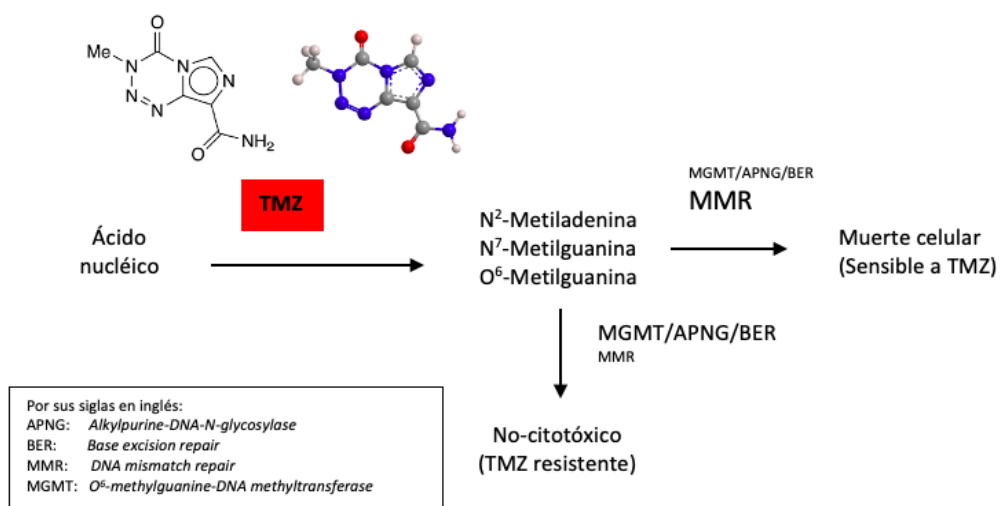
La relevancia pronóstica de los patrones de metilación del ADN específicos de tumores se ha demostrado, por ejemplo, en la leucemia linfocítica crónica y en la leucemia mielomonocítica juvenil (Weichenhan et al. 2020).

Otro ejemplo es el estado de metilación en el promotor MGMT (*Methylguanine-DNA methyltransferase*, por sus siglas en inglés) es usado para determinar si la

temozolomida, un agente alquilante del ADN, es efectivo como tratamiento contra el glioblastoma. La premisa consiste en que el MGMT es una enzima reparadora del DNA que cuando se encuentra metilada es expresada en bajos niveles, por lo que el tratamiento con temozolomida será efectivo al bloquear la reparación en las células cancerosas inducidas por el tratamiento, conduciendo a su muerte celular (Figura 7). Cuando el MGMT no está metilado, el gen es expresado y el daño por el tratamiento se limita a la habilidad de la célula para repararlo (Lee 2016).

De manera similar, la expresión específica de miRNA y moléculas de lncRNA pueden ayudar en la identificación de células tumorales resistentes a tratamientos en pacientes con quimioterapia, identificando qué pacientes necesitan un tratamiento adicional pues proveen oportunidades de proporcionar tratamientos novedosos en los tipos de cáncer en los que no existe tratamiento dirigido como el caso de glioblastoma.

Figura 7.
Mecanismo de acción y resistencia a la temozolomida.



Nota: La temozolomida (TMZ) modifica el ADN o el ARN en los sitios N7 y O6 de la guanina y el N3 de la adenina mediante la adición de grupos metilo. Los sitios metilados pueden permanecer mutados, fijarse mediante la MMR, eliminarse mediante reparación de escisión de bases (BER) o mediante la acción de una ADN glicosilasa como alquilpurina-ADN-N-glicosilasa (APNG), o desalquilar mediante la acción de una enzima desmetilante como la O6-metilguanina metiltransferasa (MGMT). Las células son sensibles a TMZ cuando MMR se expresa y se activa. Cuando se expresan las proteínas MGMT, APNG y BER, las células GBM son resistentes a TMZ.

Fuente: **Temozolomide resistance in glioblastoma multiforme by Sang Y. Lee (2016) Genes and diseases.**

4.4 Farmacogenómica

El desarrollo de fármacos durante el siglo XX se ha basado en la identificación de terapias dirigidas a toda una población. Sin embargo, en la actualidad se ha logrado dilucidar que los pacientes tienen rasgos distintivos que causan variaciones en la respuesta a la terapia por lo que se están diseñando tratamientos basados en estas características. En este rubro, la farmacogenómica tiene un papel fundamental pues se refiere al estudio y aplicación del efecto de la variación genética en la respuesta a los fármacos. Esto se vuelve muy importante para cualquier individuo que esté en tratamiento con cualquier medicación (Ashley 2016).

Teniendo actuales desafíos en el desarrollo de nuevos medicamentos, el conocimiento sobre la influencia de la variación genética en la respuesta al fármaco es relevante para predecir el resultado terapéutico. Convirtiendo a la farmacogenómica como un factor establecido en la determinación de la eficacia y seguridad de los medicamentos y permitiendo la creación de pautas de tratamiento por parte de consorcios científicos internacionales para su la aplicación clínica como lo son el Consorcio de Implementación de Farmacogenética Clínica (CPIC, por sus siglas en inglés) y el Grupo de Trabajo de Farmacogenética Holandés (DPWG, por sus siglas en inglés) que han desarrollado pautas validadas para varias interacciones entre fármacos y genes y que se encuentran disponibles como un recurso de consulta pública online. Una aplicación más general es la referente a los ajustes de dosis según el genotipo del paciente para ciertos medicamentos y clases de medicamentos que ya se han implementado en las guías clínicas relevantes(Primorac et al. 2020; Cecchin and Stocco 2020).

De hecho, el recurso financiado por los Institutos Nacionales de Salud de EU (INH, por sus siglas en inglés) llamado "*Pharmacogenomics Knowledgebase*" (PharmGKB) está dedicado a la curación y difusión de la información de farmacogenómica (PGx, por sus siglas en inglés). Las etiquetas de medicamentos que contienen información de PGx están aprobadas por agencias reguladoras mundiales como incluidas: la FDA, la EMA y la Agencia de Dispositivos Médicos y Farmacéuticos de Japón(Primorac et al. 2020).

Un ejemplo que combina la farmacogenómica y la medicina personalizada de gran importancia es el ajuste de la terapia según los polimorfismos del CYP450, la cual es una familia de enzimas responsable del metabolismo de la mayoría de los fármacos. Las variaciones en los genes altamente polimórficos que codifican algunas de las enzimas CYP determina su capacidad para metabolizar ciertos fármacos teniendo como consecuencia la existencia de diferencias en la eficacia y la seguridad en ciertos pacientes. Lo que permite identificar a los llamados "metabolizadores pobres" mediante la detección de menos alelos activos o bien alelos inactivos, lo cual puede causar sobredosis o aumentar la toxicidad. En cambio, los "metabolizadores ultrarrápidos" expresan múltiples copias de un gen particular que tendrá como consecuencia una probable falta de eficacia(Primorac et al. 2020).

También podemos tener el escenario contrario cuando analizamos el metabolismo de los profármacos. Donde los metabolizadores ultrarpidos presente niveles séricos incrementados de los componentes activos aumentando la toxicidad y los metabolizados pobres puedan ser no efectivos, por lo que los perfiles farmacocinéticos también obtenidos a través de enfoques estadísticos y bioinformáticos innovadores, tienen el potencial de explicar las fuentes predecibles de

la variabilidad entre pacientes en los efectos farmacológicos, que implementados adecuadamente conducirán a una terapia de precisión(Cecchin and Stocco 2020).

Con el objetivo anterior, es que el aprendizaje profundo o *deep learning (DL)*, el cual es un tipo de aprendizaje automático que imita la forma del aprendizaje humano, podría tomar un papel importante en la superación de los retos relativos a la implementación PGx en el EHR, como la resolución de la alta dimensionalidad de la estructura de datos de EHR, es decir reducir los efectos negativos derivados del aumento del número de variables, el ruido, la heterogeneidad, la escasez, campos de datos incompletos, error aleatorio, sesgos sistemáticos y la extracción de fenotipos clínicos relevantes, entre otros. En relación con PGx, DL podría desempeñar un papel fundamental en la predicción de la respuesta individual a los medicamentos, así como en la selección y dosificación de medicamentos, todo basado en datos analizados de grandes bases de datos(Primorac et al. 2020).

Además, los resultados del genotipo podrían resultar de crucial importancia para caracterizar mejor las diferencias interétnicas en la respuesta a los fármacos y pueden ayudar en el desarrollo de fármacos.

4.5 Exposómica

El exposoma puede ser definido como una aproximación sistemática para la adquisición de grandes cantidades de datos referentes a la exposición ambiental que tiene un individuo a lo largo de su vida y su correlación con un estado de salud o enfermedad específicos (Martin-Sanchez et al. 2020).

La exposómica se ocupa de medir exhaustivamente los eventos de exposición ambiental de una persona, desde la concepción hasta la muerte. Un objetivo es

determinar los efectos combinados de estas exposiciones mediante la evaluación de los biomarcadores y las enfermedades en las que influyen. Sin embargo, el concepto de exposoma se extiende más allá de estos factores e incluye factores sociales como socioeconómicos, calidad de vivienda, relaciones sociales, acceso a servicios y experiencia de discriminación que pueden contribuir al estrés psicológico y mala salud. (National Research Council et al. 2012)

La medicina de precisión ha reconocido la necesidad de adquirir e integrar datos genéticos, ambientales y clínicos a nivel individual para lograr una mejor comprensión de las enfermedades multifactoriales y para el desarrollo de acciones preventivas, de diagnóstico y terapéuticas adaptadas a grupo de individuos con factores de riesgo similares.

Una de las iniciativas en la exposómica fue la creación de la Asociación Nacional de Médicos Informáticos fundada en Hangzhou, China en agosto de 2017. Este grupo de trabajo ha enfocado sus esfuerzos en desarrollar aspectos informáticos relacionados con el exposoma, para apoyar en el ciclo de recolección de la información, almacenamiento, integración con datos genómicos y de fenotipo, análisis, así como la conversión en nuevo conocimiento. Es por ello, que en 2018 convocaron a un panel llamado *Exposome Informatics Working Group* donde se presentaron distintos proyectos de investigación acerca de las tecnologías digitales de recolección de información y análisis (Martin-Sanchez et al. 2020).

Uno de estos proyectos fue fundado por la Comisión Europea con el propósito de apoyar la planeación de ciudades e implementar políticas de salud pública innovadoras. Este proyecto está construido con la integración de dos componentes:

1) La recolección de información proporcionada de manera voluntaria mediante una app llamada "PulseAir", acerca de su salud, estilo de vida y patrones de exposición mediante un cuestionario y sensores de contaminación instalados en su casa.

2) El análisis de los datos, que derivó en la creación de un sistema de apoyo para la toma de decisiones, el cual está disponible para las autoridades de cuidado sanitario y los planeadores de ciudades. En cuanto a la evaluación del riesgo de enfermedades y su prevención, el proyecto se enfocó en realizar la relación entre la contaminación del aire y el desarrollo de asma, así como la relación entre la inactividad física y el riesgo de desarrollar diabetes de tipo 2.

Otro de los proyectos presentados por este grupo de trabajo, está relacionado con el clima y el dolor, pues este factor ha sido considerado como uno de los que influye mayoritariamente en la gente que vive con artritis u otras condiciones de dolor crónico. Múltiples investigaciones habían tratado de establecer esta asociación, sin embargo, los resultados siempre habían sido inconclusos por las limitaciones de los estudios, entre las cuales se incluían tamaños de muestra muy pequeñas o cortas duraciones de los seguimientos.

Para recopilar los datos se utilizó una aplicación para teléfonos inteligentes que permitió el registro diario de los síntomas más frecuentes de la artritis mediante un cuestionario. El programa "*Cloudy with chance a pain*" del Reino Unido dirigió estos cuestionarios y el estudio reclutó 14 mil pacientes que fueron interrogados a diario durante 6 meses, mientras el GPS del teléfono registraba los datos del clima local sin importar la ubicación del usuario, lo cual otorgaba una ventaja de movilidad al participante, sin que ello implicase interrumpir el seguimiento. Al final de este estudio los resultados demostraron que una humedad relativa alta, fuertes vientos o la baja

presión, están asociados con los días en los que el dolor incrementa. Este estudio no sólo demostró la importancia de la relación entre los factores climáticos y el dolor, también evidenció la utilidad del uso de dispositivos en las investigaciones de salud a gran escala (Martin-Sanchez et al. 2020).

Otra de las ventajas que otorga la tecnología es la portabilidad, esta visión está dirigida en transformar los servicios de cuidado de la salud de manera digital, pues actualmente para las condiciones crónicas, los pacientes realizan visitas clínicas innecesarias durante las etapas de remisión, en contraste con las etapas de recaída, las cuáles no son atendidas a tiempo y derivan en complicaciones u hospitalizaciones, es por ello que se pretende que mediante dispositivos portables se realicen mediciones constantes de parámetros como la frecuencia cardiaca, temperatura, presión arterial, etc que sean relacionados con los datos del EHR del paciente y que a su vez sean integrados en un modelo de predicción de riesgo de recaída de la enfermedad, permitiendo así la elaboración de un plan de cuidados y acciones necesarias, tomando en consideración el contexto de cada paciente.

Para la implementación de esta alternativa se debe promover la aceptación del usuario, explicando los beneficios económicos y de salud, los riesgos de seguridad del paciente, así como el marco regulatorio a aplicarse en la recolección y uso de la información necesaria. Este tipo de modelos ya están siendo desarrollados, como ejemplo, la Universidad de Manchester y Nueva York están desarrollando prototipos de "*wearable clinics*" para esquizofrenia, en la cuál se ha utilizado una app que resume el estatus del paciente en tiempo real. Para el caso de la esquizofrenia se realizan cuestionarios de evaluación de síntomas psicóticos, en conjunto con el monitoreo del

comportamiento mediante el GPS, así como la evaluación de riesgo en tiempo real (Martin-Sanchez et al. 2020).

Es por todo esto, que la exposómica ha demostrado su valor en la construcción del modelo de la medicina de precisión, no obstante, es necesario el desarrollo de metodologías que nos permitan realizar estimaciones de la exposición ambiental de los individuos y su relación con la salud, con la misma precisión y fiabilidad como lo haríamos con otras disciplinas como la genómica.

4.6 Metabolómica

Los metabolitos son las moléculas de bajo peso molecular que se producen a partir de los procesos bioquímicos o enzimáticos, como resultado del metabolismo celular. El metaboloma representa el conjunto de metabolitos de un sistema biológico en una determinada condición y la metabolómica es la disciplina que se encarga de caracterizarlo. La metabolómica es un campo de las ciencias ómicas que utiliza técnicas de química analítica de vanguardia en conjunto con métodos computacionales avanzados para caracterizar mezclas químicas complejas (Wishart 2016).

La caracterización del metaboloma de un órgano o tejido puede ser de gran utilidad para el diagnóstico de enfermedades o para evaluar la efectividad de un tratamiento. Los principales enfoques utilizados en metabolómica para obtener el fenotipo metabólico o *metabotype*, son la concentración absoluta o relativa de los metabolitos en una muestra, el análisis dirigido, el perfil metabólico y la toma de huellas metabólicas o *metabolic fingerprinting* (Wishart 2016).

El análisis dirigido es un enfoque cuantitativo que mide la concentración de un conjunto limitado de metabolitos conocidos. La metabolómica global o no dirigida,

intenta medir un gran conjunto de metabolitos en la muestra sin saber inicialmente qué metabolitos se espera que estén presentes en cada muestra individual. Los perfiles de metabolitos no dirigidos deben ser lo más comprensibles posibles para descubrir la red metabólica subyacente que ha sido modificada en un estado patológico. Una de las diferencias de la metabolómica frente a otras tecnologías ómicas como la genómica y sus escalas de riesgo, es que éstas son utilizadas para predecir qué va a suceder, mientras que el perfil metabólico nos indica que es lo que actualmente está ocurriendo en el organismo (Wishart 2016).

Para el caso de la huella metabólica, se considera el perfil metabólico total como un patrón particular o huella digital que caracteriza a un estado metabólico particular sin identificar de manera individual cada metabolito dentro del perfil. Por la información que se reúne en cada enfoque de la metabolómica, es que cada uno puede ser de utilidad para la medicina de precisión al comparar la presencia, ausencia, incremento o decremento de metabolitos en los distintos estadios de la enfermedad.

La metabolómica requiere de diversos instrumentos para realizar sus determinaciones, estos pueden incluir instrumentos de componente-específicos como ensayos colorimétricos para la detección de componentes redox, espectrómetros ultravioletas o fluorescentes para la detección de componentes aromáticos, espectrómetros de masa acoplados inductivamente para la detección de metales, etc. Siendo las principales técnicas analíticas en metabolómica la resonancia magnética nuclear (NMR, por sus siglas en inglés), la cromatografía de gases acoplada a espectrometría de masas (GC-MS, por sus siglas en inglés) y la cromatografía líquida acoplada a espectrometría de masas (LC-MS, por sus siglas en inglés).

La NMR se utiliza ampliamente debido a su naturaleza no destructiva y su capacidad para medir simultáneamente muchos compuestos orgánicos presentes en la muestra biológica, así como la identificación de nuevos componentes químicos, es una tecnología rápida pues sólo demora de 2 a 3 minutos por muestra, es compatible con muestras sólidas y líquidas y puede ser completamente automatizada. Esto es especialmente importante cuando el tamaño de la muestra es limitado, lo cual es común en aplicaciones biomédicas. La principal limitación de la NRM es su baja sensibilidad que permite la detección de metabolitos sólo a nivel micromolar, tiene un alto costo inicial, no puede detectar o identificar componentes inorgánicos y requiere volúmenes de muestra entre 1 y 5 mL. Pese a todo ello, la NRM se ha utilizado en gran medida para la toma de huellas metabólicas, el descubrimiento de biomarcadores y el análisis de flujo metabólico.

Por otro lado, la espectrometría de masas (MS) es otra alternativa, debido a su alta sensibilidad y selectividad, para evaluaciones cuantitativas. La utilización de MS junto con cromatografía, especialmente cromatografía líquida (LC), ha sido el estándar de oro para el análisis cuantitativo de varios compuestos. El uso de la cromatografía para el análisis de muestras clínicas complejas podría proporcionar resultados analíticos deseables, pero requiere una preparación de la muestra que suele ser laboriosa y requiere mucho tiempo. Por lo tanto, los métodos de cromatografía-EM están limitados a su uso en laboratorios clínicos (Bittremieux et al. 2018).

La alta precisión de los instrumentos Espectrometría de masas de alta resolución, (HRMS por sus siglas en inglés) utiliza espectrómetros de masas capaces de realizar mediciones de alta resolución y precisión de masas, por lo cual permiten una mejor caracterización de metabolitos desconocidos mediante la asignación de fórmula

elemental y también permiten la identificación de aductos con alta precisión, de la misma manera el incremento de la sensibilidad de las nuevas metodologías permiten utilizar volúmenes de muestras más pequeños, lo cual cobra relevancia pues muchas de las muestras disponibles en biobancos se encuentran en cantidades limitadas.

Por otro lado, otras tecnologías como la GC-MS tienen otro tipo de ventajas como un menor costo inicial, cuantitatividad con calibración, requiere un volumen de muestra menor entre 1 y 2 mL, alta sensibilidad, la detección de componentes orgánicos y moléculas inorgánicas y tiene reproducibilidad de separación. Puede ser automatizada en su mayoría y es compatible con gases y líquidos.

A diferencia de la NMR, la técnica la GC-MS si es destructiva. La muestra sólo puede ser procesada una vez y no es recuperable, requiere derivación y separación de muestra, es lenta pues toma de 20 a 40 min por muestra, no es compatible con sólidos, es difícil realizar la identificación de nuevos componentes.

En el caso de la última tecnología LC-MS, sus ventajas principales son su alta sensibilidad, también puede detectar la mayoría de las moléculas orgánicas e inorgánicas. Requiere de tamaños de muestra muy pequeños, orden de 10 a 100 microlitros, se puede realizar sin separación. Puede ser automatizada y es compatible con sólidos y líquidos. Como desventaja y al igual que la GC-MS, es una técnica destructiva, con un alto costo inicial, alto tiempo de procesamiento, entre 15 a 40 minutos por muestra. No es compatible con gases, se dificulta la identificación de nuevos componentes, así como una corta vida del instrumental utilizado (menos de 9 años).

La obtención de imágenes de metabolitos implica la detección y visualización *in vivo* o *in vitro* de metabolitos en tejidos mediante: RMN, espectroscopia de resonancia

magnética (MRS), tomografía por emisión de positrones (PET, por sus siglas en inglés), desorción / ionización láser asistida por matriz (MALDI) -MS, (SIMS) o técnicas de desorción MS de iones secundarios (DESI-MS), ionización por electropulverización MS. La MRS y la PET son particularmente atractivas porque permiten obtener imágenes metabólicas no invasivas para el diagnóstico médico y el fenotipado metabólico.

Las mejoras en las secuencias de pulsos de RMN, la intensidad del campo magnético, los protocolos de etiquetado químico y la velocidad de procesamiento de la computadora están permitiendo que se detecte un mayor número de compuestos a concentraciones más bajas. En algunos casos, se pueden identificar y cuantificar parcialmente hasta 20 metabolitos diferentes en determinados tejidos. Estas mejoras están conduciendo a diagnósticos mucho más precisos y terapias personalizadas.

Una contribución importante de la metabolómica en el campo de la medicina de precisión tiene que ver con el desarrollo de nuevos fármacos, rompiendo con el paradigma tradicional de la búsqueda de una causa genética para el desarrollo de las enfermedades, clonación de genes identificados, purificación de proteínas diana y cribado de alto rendimiento para identificar posibles pistas de fármacos, que luego se optimizan y prueban en modelos animales y, finalmente, en ensayos en humanos.

Sin embargo, este modelo clásico no ha sido totalmente exitoso, algunos de los inconvenientes asociados con el modelo existente es que son relativamente pocas enfermedades las que tienen una base genética sólida, menos del 10 %. Muchas enfermedades son el resultado de exposiciones a condiciones externas (es decir, donde existe intervención del exposoma) y no todos los genes que causan

enfermedades son susceptibles de cribado de alto rendimiento o se pueden dirigir con fármacos.

Es por ello, que la metabolómica puede ofrecer una ruta mucho más rentable y productiva para el descubrimiento, prueba y desarrollo de fármacos dado que muchas de las enfermedades actualmente predominantes como enfermedades cardíacas, diabetes, obesidad, hipertensión, depresión o la enfermedad inflamatoria intestinal, tienen una fuerte base metabólica o una clara causa metabólica.

Por otra parte, se ha descubierto que algunas enfermedades crónicas (incluido el autismo, la esquizofrenia, el asma, el cáncer y la enfermedad de Alzheimer) tienen causas metabólicas de asociación inesperadas o no apreciadas anteriormente. El metabolismo desregulado es uno de los sellos distintivos del cáncer, por ejemplo, el efecto Warburg que es el cambio de un metabolismo celular aerobio a uno anaerobio, se ha considerado como un factor que permite la transformación neoplásica de la célula fenómeno conocido como hipótesis de Warburg. Recientemente, el descubrimiento de alteraciones del metabolismo de glucosa y lípidos en pacientes con enfermedad de Alzheimer y pre-Alzheimer, ha dado lugar a una serie de nuevas teorías sobre el desarrollo y la etiología de la enfermedad. Estos hallazgos en conjunto, han llevado a muchos a considerar la posibilidad de cambiar el nombre de la enfermedad de Alzheimer como "diabetes tipo 3" (Wishart 2016).

Muchos de los fármacos más eficaces son inhibidores enzimáticos, y muchos de ellos son inspirados en metabolitos o antimetabolitos. En general, si un metabolito o un conjunto de metabolitos se identifica como causal, significa que el blanco del fármaco es una enzima y se conoce su vía biosintética.

El descubrimiento de una base metabólica para una enfermedad conduce a una solución terapéutica simple, estabilizar los niveles de un metabolito que está demasiado alto o bajo.

Por otra parte, la diversidad de aplicaciones de la metabolómica parte del hecho de que puede ser utilizada para analizar un amplio rango de sustratos, incluyendo sólidos (como tejidos, residuos biológicos), líquidos (biofluidos), gases (aliento, humo, aromas), lo cual permite incluso realizar análisis in vivo.

En conclusión, la metabolómica es un poderoso enfoque de las tecnologías ómicas que ya encontró su posición en aplicaciones médicas y la medicina de precisión, por lo cual se deben desarrollar bases de datos médicos que incorporen datos de metabolómica, que su vez se puedan convertir en un fenotipo metabólico para diversas enfermedades.

4.7 Proteómica

La función de las proteínas es representativa de los procesos biológicos debido a que controlan y catalizan muchas de las reacciones bioquímicas. Una variedad de factores, incluido el nivel de transcripción, la eficiencia de traducción, el transporte de proteínas, la síntesis de proteínas y el recambio, afectan la cantidad de proteínas traducidas por molécula de transcripción.

El estudio del proteoma se ha convertido en una herramienta informativa muy valiosa pues describe la composición proteica de una célula en una condición definida ya sea en un estado de salud o enfermedad. La proteómica dirigida consiste en utilizar métodos basados en espectrometría de masas y resonancia magnética nuclear para realizar mediciones de muestras de manera cuantitativa, reproducible, imparcial,

sensible, precisa y a gran escala, lo cual permite realizar investigaciones sistemáticas acerca del comportamiento dinámico de los sistemas biológicos en diferentes condiciones a lo largo del tiempo o bien la comparación significativa de los patrones de proteoma obtenidos de grandes cohortes de muestras en estudios clínicos.

El análisis de grandes cohortes de muestras es importantes en los estudios clínicos traslacionales, debido al alto grado de variabilidad biológica en la población y a que los eventos específicos de una enfermedad sólo pueden determinarse con base en análisis estadísticos que tengan la potencia suficiente. Además de determinar la cantidad de proteínas, la proteómica también nos permite el estudio sistemático de las modificaciones postraduccionales de las proteínas (PTM, por sus siglas en inglés), la formación de complejos, la localización de proteínas, particularmente en proyectos de traducción, ya que las enfermedades pueden afectar el proteoma en diferentes niveles.

Los mecanismos antes mencionados controlan eventos de señalización celular, los cuales están involucrados en la adaptación celular a distintas condiciones, que tienen que ver con la interacción, transporte o localización. Así mismo las alteraciones de estos procesos debido a una enfermedad pueden producir perfiles postraduccionales que sirven como indicativos del fenotipo de la enfermedad y por consiguiente son candidatos a biomarcadores que permitan llevar a cabo intervenciones terapéuticas. Los biomarcadores proteicos usualmente modifican su abundancia en paralelo con el fenotipo de la enfermedad de interés.

El método de elección actual para identificar y cuantificar muestras de proteínas complejas es la cromatografía líquida (LC, por sus siglas en inglés) seguida de la espectrometría de masas (MS, por sus siglas en inglés). La importancia de estas

técnicas se ejemplifica por su uso en iniciativas de investigación a gran escala, como los dos intentos recientes de proporcionar un borrador del proteoma humano el Proyecto del Proteoma Humano (HPP, por sus siglas en inglés) en curso por la Organización del Proteoma Humano (HUPO, por sus siglas en inglés) (Bittremieux et al. 2018).

Un ejemplo es el descubrimiento de que la abundancia de la proteína MeCP2, que está directamente relacionada con la severidad de los síntomas del síndrome de Retts, la cuál es una enfermedad neurodegenerativa, y nos ayuda a ejemplificar la utilidad de la proteómica en la medicina de precisión, pues se determinó que cambios en la abundancia de la proteína MeCP2 pueden derivar en distintos trastornos con síntomas traslapados. Esto permitió proponer una intervención terapéutica que involucra la inhibición farmacológica de la proteína que modula la estabilidad y función de MeCP2 (Uzozie and Aebersold 2018).

4.8 Integración de datos obtenidos mediante ciencia básica

Debido a una cuestión histórica, por mucho tiempo se ha utilizado de manera indistinta el término de medicina de precisión con medicina genómica. Sin embargo, para que realmente la medicina de precisión pueda aplicarse es necesario hacer uso de todas las tecnologías ómicas, pues tal y como se ha descrito previamente, cada una de ellas proporcionan información relevante y complementaria para obtener una imagen completa de los procesos biológicos implicados en el desarrollo de la enfermedad.

Es necesario desarrollar proyectos que integren la información de diferentes moléculas tales como ARN, proteínas y metabolitos, utilizando la información conocida de mecanismos y vías moleculares para generar nuevo conocimiento. Uno de los proyectos más relevantes es *Encyclopedia of DNA Elements* (ENCODE) que

vincula la información acerca de la regulación de la expresión genética con la variación de secuencia del ADN y la epigenética en varios tipos de células. (Olivier et al. 2019)

Otro enfoque para abordar la integración de la información está basado en buscar correlaciones entre diferentes tipos de datos, sin considerar la información referente a las vías o redes de información celular. Esto permite identificar correlaciones de moléculas o metabolitos que están siendo transformados durante el desarrollo de la enfermedad., asu vez es posible descubrir nuevas moléculas y vías esenciales para el desarrollo de la patología. Sin embargo, implica un reto informático y estadístico pues muchas plataformas analizan y cuantifican exhaustivamente las moléculas de interés, mientras otras solo lo ejecutan con las moléculas más abundantes. Realizar la integración de datos provenientes de varias tecnologías ómicas requerirá de estrictos modelos estadísticos que ayuden en la interpretación de los grupos multi ómicos.

Por otra parte, la iniciativa *Trans-Omics for Precision Medicine* (TOPMED) del (*National Heart, Lung and Blood Institute* (NHLBI) es un ejemplo de un esfuerzo a gran escala que está comenzando a generar datos ómicos más completos para una amplia gama de estudios de cohorte, basados en muestras ya existentes, el objetivo de TOPMED es mejorar la comprensión de enfermedades específicas mediante la recolección de datos como secuenciación del genoma completo aunado a otros datos ómicos (Olivier et al. 2019).

Actualmente los resultados esta iniciativa TOPMED dependen de la disponibilidad de muestras de tejido adecuadas y material de biopsia recolectado que permita el análisis del transcriptoma, epigenoma, proteoma y metaboloma. Hasta hace poco, las

muestras de biopsia se recolectaban principalmente para el análisis de transcriptomas o el descubrimiento de mutaciones y, en consecuencia, las muestras se recolectan en tampones que estabilizan el ARN y el ADN, por lo que el material de la muestra no es adecuado para el perfil de proteínas u otros metabolitos. Los estudios futuros deberán abordar esta problemática para que podamos generar los datos multi ómicos necesarios y completos(Olivier et al. 2019).

Otro ejemplo, comentado por Denny and Collins 2021, que ha facilitado la fusión de datos provenientes de múltiples cohortes y tecnologías ómicas es el realizado por la *Global Alliance for Genomics and Health (GA4GH)*, la cual ha desarrollado modelos de datos y formatos de archivos comunes para facilitar la colaboración entre expertos de cada cohorte y la interoperabilidad de los datos. Esto permite y facilita a médicos, biólogos e investigadores construir modelos más completos de las enfermedades al crear un estándar que fomenta la sinergia entre ellos, las organizaciones y los sistemas que componen el modelo de la medicina de precisión. La rápida adopción de los estándares de intercambio de información podría mejorar la velocidad de un intercambio completo y preciso de información relevante para la investigación y atención médica (Denny and Collins 2021).

Uno de los esfuerzos más ambiciosos para evaluar las oportunidades y desafíos en el uso de datos multiómicos integrados para la evaluación y predicción de la salud es el proyecto de perfiles ómicos personales integrados *Integrated Personal Omics Profiling (iPOP,)* (Olivier et al. 2019) . Inicialmente se propuso para elaborar los perfiles ómicos de un solo individuo a lo largo del tiempo (longitudinales) y sus pruebas clínicas. El proyecto también ha explorado el impacto de los cambios de estilo de vida en la salud a largo plazo para evaluar los parámetros que son útiles en la toma

de decisiones. El último estudio conducido por esta iniciativa dio seguimiento a 109 personas en riesgo de diabetes tipo II durante aproximadamente 3 años y utilizó evaluaciones clínicas, con dispositivos portátiles y perfiles multiómicos para derivar modelos predictivos de resultados de salud a largo plazo. El análisis destaca cómo los datos resultantes permiten hacer predicciones y tomar decisiones de tratamiento en una amplia gama de especialidades clínicas, incluida la salud cardiovascular, las enfermedades infecciosas y la oncología. Si bien este estudio tiene limitaciones y no proporciona necesariamente un plan claro para los estudios multiómicos y las aplicaciones clínicas, es el primero de su tipo en evaluar el papel potencial de los datos multiómicos en las predicciones de resultados de salud de manera integrada (Quezada et al. 2017).

5 Investigación aplicada

5.1 Biomarcadores

Los biomarcadores son definidos por la Organización Mundial de la Salud (OMS) como “Cualquier sustancia, estructura o proceso que puede ser medido en el cuerpo o sus productos y que tiene influencia o predice la incidencia o resultado de una enfermedad”.

Los biomarcadores pueden ser clasificados en cuatro tipos:

- a) **Diagnóstico.** Usados para determinar la presencia de una enfermedad específica.
- b) **Pronóstico.** Ayudan a trazar el curso probable de la enfermedad.
- c) **Predictivos.** Indican la probable respuesta a un tratamiento en particular
- d) **Predisposición a la enfermedad.** Indican el riesgo de desarrollar una enfermedad.

En la medicina de precisión existen diversos ejemplos de éxito del uso de biomarcadores como: el tipo de sangre usado como referente para la transfusiones sanguíneas exitosas (Predictivo), el tamizaje metabólico del recién nacido para la detección temprana de enfermedades metabólicas (Predisposición a una enfermedad), el análisis sérico del antígeno prostático para la detección temprana de cáncer de próstata (Predisposición a una enfermedad), sobreexpresión o amplificación del receptor HER-2 en células de cáncer de mamá como un predictor de la respuesta a anticuerpos monoclonales como trastuzumab o pertuzumab (Predictivo), entre otros (Quezada et al. 2017).

El avance de las nuevas tecnologías ómicas ha sido fundamental en el descubrimiento de nuevos biomarcadores, debido a que es posible analizar moléculas como el ADN, los péptidos o los metabolitos de manera simultánea y exhaustiva en individuos con condiciones específicas. Estos datos son analizados para determinar cuáles de estas moléculas pueden ser usadas como biomarcadores asociados con la incidencia de una enfermedad, y para proporcionar un pronóstico o determinar la respuesta ante una intervención terapéutica definida. Esto propicia y facilita un tratamiento individualizado para el paciente y el monitoreo del progreso con la terapia proporcionada. Uno de los ejemplos más notables son los biomarcadores derivados de la transcriptómica basados en el perfil de expresión genético tumoral, Mamaprint, OncotypeDX y PAM50 que abordan el riesgo de recurrencia de cáncer de mama.

El desarrollo de un biomarcador para uso clínico consta de 4 fases (Figura 8) (Quezada et al. 2017)

1. **Descubrimiento.** Se analizan los bioespecímenes candidatos a usarse como biomarcadores, éstos pueden provenir de líneas celulares, modelos animales,

biopsias provenientes de cohortes, muestras de pacientes enrolados en estudios clínicos en curso, muestras almacenadas de estudios prospectivos finalizados y biobancos. Estos datos serán analizados desde el contexto de la información disponible.

Idealmente las muestras deberían originarse de amplios estudios caso-control prospectivos con alto control de las características clínicas involucradas, pero cuando esta condición no es posible se deberán controlar los criterios de inclusión en los datos a analizar, ya que, al incluir muestras heterogéneas de grupos poco definidos, los resultados tendrán poco valor clínico, pues las muestras no reflejarán la población de pacientes en donde los biomarcadores quieren ser usados.

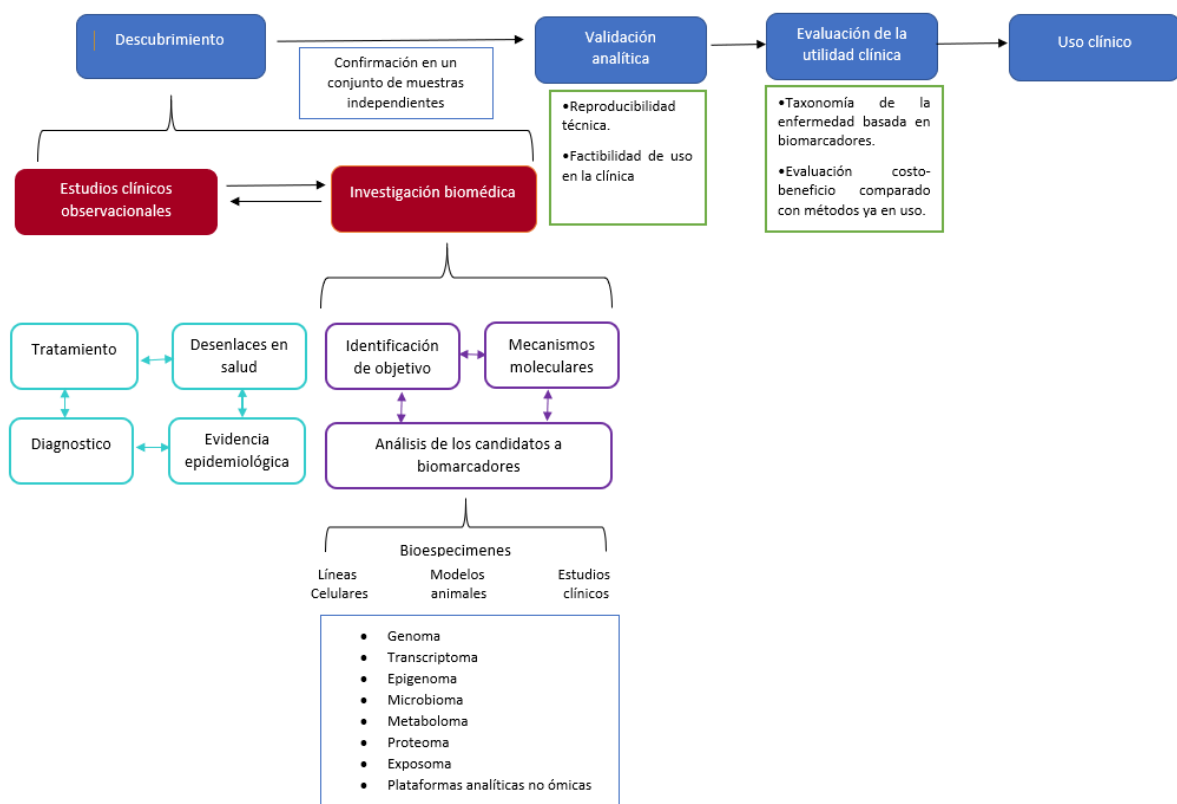
El manejo de las muestras también es parte fundamental en el descubrimiento de los biomarcadores, siendo necesario un protocolo de recolección, almacenamiento, procesamiento y caracterización que permita su validación mediante métodos analíticos reproducibles.

2. Validación analítica. Una vez que se identifiquen los candidatos a biomarcadores, se deberá evaluar su utilidad mediante análisis disponibles en el laboratorio clínico, como FISH (*Fluoresce in situ Hybridization*), RT-PCR (*Polymerase Chain Reaction*), PCR, HPLC o algún ensayo basado en inmunoafinidad. Estos análisis tienen por objetivo la validación analítica del rango de detección y la reproducibilidad. Se han desarrollado biomarcadores multinivel que incluyen el análisis de diversos tipos de moléculas, los cuales cobran importancia cuando una condición patológica está determinada por varios tipos de moléculas como ADN, proteínas, ARN o metabolitos.

- Evaluación de la utilidad clínica.** Para confirmar la utilidad clínica de un biomarcador de diagnóstico o predictivo de resultado clínico se pueden realizar estudios clínicos prospectivos en donde el marcador dirija el tratamiento del paciente o en estudios prospectivos/retrospectivos donde se analicen muestras almacenadas, así como las presentes en biobancos.
- Uso clínico.** Después de demostrar la utilidad clínica, el biomarcador debe obtener la aprobación regulatoria, ser comercializado e incorporado a las guías de práctica clínica.

Figura 8.

Proceso de desarrollo de un biomarcador



Fuente: *Omics-based biomarkers: current status and potential use in the clinic*, (p. 219-226) por Hector Quezada *et al.*, (2017), Bol Med Hosp Infant Mex

Una de las aplicaciones más exitosas de los biomarcadores ha sido en el campo de la oncología. En el caso de las neoplasias, las tecnologías ómicas nos permiten estudiar los factores genómicos de riesgo y los perfiles de mutación asociados. En el caso de los primeros se realiza un estudio de asociación del genoma completo *Genome-Wide Association Studies* (GWAS) , comparando la distribución de las variantes más comunes a lo largo del genoma en grandes cantidades de casos y controles, para los perfiles de mutaciones en los tumores se estudian a través de la secuenciación de próxima generación (NGS, por sus siglas en inglés) del genoma, el exoma, el transcriptoma o un panel de genes que se sabe que están involucrados en una neoplasia particular.(Quezada et al. 2017).

Una de las características deseables en los biomarcadores es que deben ser analizados en bioespecímenes no invasivos como sangre, plasma, suero, orina, saliva o heces. Los metabolitos en biopsias líquidas como suero u orina pueden reflejar el estado bioquímico del paciente, el cual será el resultado de interacciones entre medicamentos, expresión genética, proteínas, edad, microbiota, ambiente y enfermedades. Estos perfiles metabólicos pueden ser una fuente promisoría de biomarcadores no invasivos. Un ejemplo de esto se realizó en un estudio utilizando el plasma de 34 pacientes HER-2 positivos con cáncer de mama con diferentes respuestas a la terapia neoadyuvante con trastuzumab-paclitaxel, revelando mediante el perfil metabólico que los pacientes con altas concentraciones de espermidina y baja concentración de triptófano tenían altas probabilidades de presentar una respuesta completa, comparados con los pacientes con distinto balance de estos metabolitos. Por lo que la tasa de triptófano/espermidina fue usada para establecer un valor umbral y construir una curva de características de receptor-

operador (ROC, por sus siglas en inglés) con sensibilidad del 90% y especificidad del 87% (Muraro et al. 2015; Miolo et al. 2016).

Las tecnologías ómicas nos permiten caracterizar bioespecímenes a nivel molecular, aunque la complejidad de las enfermedades representa aún un reto para estas tecnologías. Las dificultades están vinculadas en cuanto a la reproducibilidad y alta tasa de falsos positivos, por lo que, todos los candidatos a biomarcadores descritos en estudios preclínicos deberán ser validados en estudios prospectivos bien diseñados para darles una validación clínica permitiendo una mejor evaluación y estratificación de los pacientes.

5.2 Reposicionamiento de fármacos

El reposicionamiento de fármacos es una estrategia para identificar nuevos blancos terapéuticos para medicamentos ya aprobados o en investigación. Este enfoque ofrece varias ventajas sobre el desarrollo de un fármaco completamente nuevo para una determinada indicación. Primero el riesgo de fracaso es menor debido a que ya se ha demostrado que el fármaco es seguro en modelos preclínicos y en humanos si se han completado los ensayos en etapa inicial. En segundo lugar, se reduce el tiempo para el desarrollo y el monto de la inversión económica porque la mayoría de las pruebas preclínicas, la evaluación de la seguridad y, en algunos casos, el desarrollo de la formulación ya se habrán completado (Pushpakom et al. 2019)

Un ejemplo de esta estrategia es el que se refiere al medicamento Ivacaftor desarrollado para pacientes con fibrosis quística (FQ). Esta es una enfermedad crónica y progresiva, de etiología autosomal recesiva que afecta en promedio a 70,000 personas en el mundo y es causada por variantes en el gen que codifica la proteína reguladora de la conductancia transmembrana de la fibrosis quística (CFTR,

Cystic fibrosis transmembrane conductance regulator, por sus siglas en inglés). La proteína producto de este gen es un canal iónico epitelial localizado en la superficie de la célula donde se regula el tránsito celular de cloro. Las mutaciones del gen CFTR causa una regulación anormal de sal y agua, afectando principalmente a los pulmones, páncreas y glándulas sudoríparas. La principal complicación de la enfermedad es la deshidratación de las secreciones y la producción de un moco espeso responsable del desarrollo de infecciones, convirtiéndose en el mayor marcador pronóstico de la enfermedad.

La FQ ha sido clasificada en 6 subtipos según la mutación que afecta el gen CFTR y que se traducen en el defecto funcional que causan. Ivacaftor, inicialmente fue dirigido a la población con la subclasificación III, que presentaban la mutación G551D, la cual representa 5 % de los pacientes. Esta mutación genera una regulación defectuosa en el canal y un transporte inadecuado de cloruros, por lo Ivacaftor fue diseñado para incrementar el tiempo de activación de los canales CFTR en la superficie de la célula

Sin embargo, en el genotipo F508 el cual es el más común de la enfermedad y representa 85 % de la población. los canales mutantes no alcanzan la superficie celular por un mal plegamiento de la proteína representando una menor respuesta terapéutica ante Ivacaftor, pues la pequeña cantidad de CFTR que alcanza la superficie celular tiene una probabilidad baja de apertura del canal (apertura del canal defectuoso). Lumacaftor es un corrector de CFTR que actúa directamente en F508del-CFTR para mejorar su procesamiento y transporte celular, incrementando de este modo la cantidad de CFTR funcional en la superficie celular. Ivacaftor como potenciador de la proteína CFTR facilitará el aumento del transporte de cloruro al potenciar la probabilidad de apertura (activación) del canal de la proteína CFTR en la

superficie celular. No se conocen los mecanismos exactos por los que lumacaftor mejora el procesamiento y el transporte celular de CFTR, pero en conjunto con el uso de Ivacaftor ha logrado beneficiar a un gran porcentaje de pacientes con FQ.

Este es un ejemplo de reposicionamiento de un fármaco con un indicación terapéutica inicial para una subpoblación de pacientes de FQ muy específica y que a través de la correcta estratificación o subtipificación de la enfermedad pudo posicionarse como una opción farmacológica viable, en conjunto de otro fármaco en otros subtipo de la misma enfermedad.

6 Biobancos

La implementación de la medicina de precisión necesita que el sistema de salud cumpla con requerimientos de infraestructura para la construcción de biobancos que permitan el almacenamiento de muestras biológicas y la generación de datos precisos tanto en el ámbito de la investigación como en el ambiente clínico. La investigación mediante biobancos es esencial para dilucidar la etiología de enfermedades complejas y la base molecular de los subtipos de enfermedades, pues los biobancos juegan un papel importante en la identificación de nuevos biomarcadores.

El término “biobanco” es definido como una colección de muestras biológicas humanas e información asociada, que se encuentran organizadas de una manera sistemática para propósitos de investigación. Los biobancos cubren un enfoque multidisciplinario de la salud humana que combina enfoques biológicos y médicos, así como tecnologías bioinformáticas, computación y modelado. La importancia de los biobancos ha aumentado durante la última década debido a su variedad y capacidad, pues van desde pequeñas colecciones de muestras hasta repositorios nacionales o internacionales a gran escala.

La historia del biobanco surge hace aproximadamente 30 años. Los primeros biobancos fueron depósitos de muestras e información recolectadas al azar; datos asociados con bioespecímenes almacenados que han aumentado en el tiempo en complejidad desde lo básico, como la fecha de recolección y el diagnóstico, hasta extensos conjuntos de información que abarcan muchos aspectos del fenotipo del participante o del paciente, que ahora se extienden rápidamente con información genética, proteómica y otra información ómica

Existen biobancos generales que pueden admitir una variedad de estudios, incluidos estudios transversales de correlaciones genotipo-fenotipo, estudios de casos y controles que usan un biobanco para casos y/o controles, y estudios de cohortes que usan datos de referencia y de seguimiento en un biobanco para vincular la variación genética con los resultados de salud (Kinkorová 2015) .

La actividad de “*biobanking*”, se refiere a la creación de estas colecciones de bioespecímenes, lo cual incluye numerosos procesos como el enrolamiento de pacientes, así como la recolección, procesamiento, almacenamiento y distribución de bioespecímenes para su uso en estudio. Anteriormente, dependiendo de las preguntas de investigación, se determinaban las variables de los biobancos como el tamaño de la cohorte, los tipos de biospecimen (tejido biopsiado, sangre, orina, ADN, etc), su estado de conservación, (frescos, congelados, fijados con parafina [FFPE]), y los métodos de almacenamiento (congelador de nitrógeno líquido, congelador mecánico ultrabajo). Las muestras que formaran parte del biobanco son recolectadas bajo condiciones estandarizadas para garantizar su reproducibilidad y la preservación de compuestos (por ejemplo, metabolitos) que pudieran ser críticos.

Actualmente existe una creciente demanda para implementar biobancos humanos, pues son la base para proporcionar cohortes de bio especímenes que permitan el descubrimiento de biomarcadores útiles para la medicina de precisión. Los biobancos deben garantizar que los especímenes biológicos tienen la calidad suficiente para que cualquier observación pueda ser atribuida a las características inherentes a la bioespecie y no a variables preanalíticas como errores de observación derivados de la manipulación, el procesamiento o el almacenamiento del bioespecimen.

De manera general hay dos tipos de biobancos, los que están destinados para laboratorios de investigación y los biobancos clínicos. Aunque a la fecha tienen algunas diferencias significativas en sus procesos e intereses a largo plazo, el avance de la implementación de la medicina de precisión está promoviendo la integración de estos dos enfoques para garantizar la calidad de los bioespecímenes y la eliminación de la variabilidad preanalítica.

Para garantizar la calidad de las muestras biológicas se debe realizar una estandarización de los procesos centrada en la adopción de mejores prácticas y el uso de procedimientos operativos estándar, y buscar la estandarización externa en relación con otros biobancos a través de un programa de certificación o acreditación. Actualmente existen tres organizaciones que han desarrollado estándares y programas de verificación externos, lo cuales son específicamente para biobancos enfocados a la investigación, pero podría ser adaptable a los biobancos clínicos

- a) *The Canadian Tissue Repository Network (CTRNet).*
- b) *The College of American Pathologist (CAP)*
- c) *The International Organization for Standardization (ISO) 20387.*

La norma del CTRNet, a través de su programa de certificación está más integrada con los requisitos de educación y formación del personal de los biobancos, por lo que es recomendable que esta sea el estándar inicial por parte de los biobancos clínicos. Sin embargo, a medida que la medicina de precisión se integre plenamente en la prestación de la atención clínica, es probable que los biobancos clínicos tengan que adoptar también las normas de biobancos CAP e ISO, más centradas en el laboratorio.

Otro ejemplo de esta práctica es el *Biobank consortium EuroBioBank* (EBB), la cual fue la primera red operativa de biobancos en Europa en proporcionar muestras de ADN, células y tejidos humanos para la investigación de enfermedades raras. Sin embargo, La infraestructura más importante y más grande de Europa e incluso a nivel mundial es la *Biobanking and BioMolecular resources Research Infrastructure—European Research Infrastructure Consortium* (BBMRI-ERIC). La misión de la BBMRI-ERIC es aumentar la eficacia y la excelencia de la investigación biomédica europea facilitando el acceso a recursos biológicos relevantes para la salud humana y las enfermedades con una calidad definida.

Una serie de consideraciones son fundamentales para mantener activos y dinámicos los biobancos y la investigación que respaldan. La necesidad de inversiones a largo plazo en recursos biológicos es claramente evidente cuando se necesitan datos longitudinales. Las recopilaciones prospectivas de datos y muestras de individuos asintomáticos permitirán la identificación preventiva y desarrollo subclínicos de desarrollo de las enfermedades e identificarán biomarcadores de pronóstico y diagnóstico y dianas farmacológicas (Kinkorová 2015)

7 Historias clínicas electrónicas

Los expedientes clínicos electrónicos son la herramienta que nos proporciona información referente al estilo de vida de los pacientes o exposoma y que puede ser considerada como factores de riesgo.

La necesidad del acceso y análisis de la información contenida en los EHR está justificada desde diversos puntos de vista. Por ejemplo, en el caso del diseño de los estudios prospectivos tradicionales, éstos se encuentran limitados por la medición de los parámetros establecidos durante el periodo del estudio, como si se tratara de un fenotipo estático. Sin embargo, las enfermedades son dinámicas y tienen componentes que cambian o se presentan a lo largo del curso de la enfermedad y que puede llegar a ser excluidos en este tipo de estudios. Es aquí donde los EHR se convierten en una fuente útil para el análisis de las características no previstas de manera longitudinal, pues contienen otros datos obtenidos durante las actividades de la rutina clínica a lo largo de la vida del paciente (Glicksberg, Johnson, and Dudley 2018).

A diferencia de los estudios genómicos donde las poblaciones implicadas no necesariamente logran representar a la población en general, los estudios observacionales nos proporcionan “*Real world data*” o datos del mundo real, los cuales son los datos obtenidos por cualquier metodología no intervencionista que se recopilan prospectiva y retrospectivamente de las observaciones de la práctica clínica habitual. Estos pueden provenir de varias fuentes siendo una de las principales el uso de los datos incluidos en los EHR, con lo cual tendríamos la misma calidad de información que la proporcionada por un estudio clínico convencional. (Gómez 2019)

Los estudios observacionales presentan la ventaja de que la información para su análisis es recolectada durante el proceso rutinario del cuidado de la salud. Esta recolección forma parte de los fundamentos de la práctica de la MP y se puede lograr a través de los datos contenidos en el EHR, pues los grandes centros académicos y los sistemas de salud cuentan con registros de millones de pacientes, los cuales son imposibles de reunir en un estudio de cohorte por un investigador de manera individual, y los EHR son la fuente de información más representativa de la práctica clínica. Como ejemplo, 44 % de las revisiones de Cochrane concluyeron que “la evidencia de los estudios clínicos presentados es insuficiente para la práctica clínica” derivado de que las poblaciones analizadas son poblaciones pequeñas y homogéneas y por tanto, no son generalizables para la mayor parte de la población objeto de estudio (Gómez 2019).

En este mismo sentido, se requiere un diseño adecuado que tenga en cuentas las limitaciones de los estudios observacionales como sesgos de confusión, selección e información. Para garantizar el rigor científico necesario se han creado las guías RECORD (*Reporting of studies Conducted using Observational Routinely-Collected Health Data*), que consideran los problemas en la investigación utilizando datos de salud que fueron recopilados de manera rutinaria.

Mediante la colección sistemática de la información del EHR se puede obtener información de fenotipos y datos para el diseño de nuevos estudios que no están disponibles en el actual acervo de investigación. Por ejemplo, un estudio demostró que sus participantes en promedio tenían más de 190 notas clínicas, 14 estudios radiológicos y más de 700 estudios de laboratorio en aproximadamente 8 años de seguimiento. Con el análisis de todos estos datos en miles de personas se puede

realizar el descubrimiento específico de endotipos o fenotipos emergentes en las enfermedades, por lo que la práctica de la medicina de precisión requerirá de acceso completo a los datos del EHR (Denny and Collins 2021). Aunque los datos en los EHR están limitados por las visitas de tipo de los pacientes, cuando estos se relacionan con los datos derivados de tecnologías ómicas, es posible realizar un análisis completo del estado de salud.

En Estados Unidos actualmente se lleva a cabo el proyecto de expedientes médicos electrónicos y genómica (eMERGE, por sus siglas en inglés) que ha podido confirmar que las asociaciones fenotipo-genotipo son igualmente válidas que las obtenidas mediante investigaciones de cohorte calificadas. Este proyecto ha puesto de manifiesto, la necesidad de que toda la red de información esté disponible para poder establecer la asociación genómica, con otros elementos para establecer fenotipos de calidad derivados de esta información tales como codificación para las enfermedades, resultados de estudios de laboratorio, historial de medicación y el procesamiento de las notas del médico (National Research Council et al. 2012).

Es pertinente hacer la mención que los EHR actuales requieren de un proceso de estandarización para que sean útiles a una escala poblacional. Esto debido a que algunos datos no se encuentran estructurados de manera homogénea como los informes narrativos y los datos de imágenes (Denny and Collins 2021), aunado a que la estructura de los EHR son el resultado de años de integración fortuita entre varias unidades clínicas, proveedores de software, sistemas informáticos, etc., los cuales no cuentan con un estándar de calidad establecido, y es posible que presenten errores sustanciales, redundancia o ausencia de información. Estos problemas de interoperabilidad pueden superarse a partir de una estructura

normalizada, como los esfuerzos realizados por el *Observational Health Data Sciences and Informatics Consortium* (Glicksberg, Johnson, and Dudley 2018)

La implementación de registros médicos electrónicos se constituye entonces en una herramienta valiosa para medir la prevalencia del fenotipo y el resultado entre los pacientes tratados y no tratados, posicionando a los centros que los posean como redes de aprendizaje (Seymour et al. 2017). Otra utilidad del uso de los EHR como herramienta es que pueden proveer a los clínicos de soporte en la toma de decisiones clínicas (CDS, *Clinical-decision support*) al enlazar y vincular resultados de una prueba genética con recursos electrónicos como websites o bases de datos (Aronson and Rehm 2015). En la tabla 2 se mencionan y comentan de forma introductoria algunos casos de uso de los CDS.

Tabla 2.

Ejemplos de uso de los CDS en la medicina de precisión.

Tipos de casos de uso	
Caso de uso	Comentarios
Mostrar resultados a los médicos	El enfoque inicial es la visualización genérica de los resultados de la farmacogenómica; las visualizaciones específicas de la enfermedad se considerarán más adelante.
Incluir pruebas genéticas en conjuntos de pedidos.	Implica la inclusión de pruebas genéticas en conjuntos de orden estándar sin comprobar si la prueba se ha realizado antes.
CDS para identificar si se debe solicitar una prueba (alerta previa a la prueba *)	Implica ejecutar un algoritmo CDS cuando se solicita un medicamento para verificar si las variantes farmacogenómicas relacionadas ya se han evaluado y advertir al médico si no lo han hecho.
CDS para identificar cuando un pedido de medicamentos no coincide con el resultado de una prueba (alerta posterior al pedido *)	Implica ejecutar un algoritmo CDS en respuesta a una orden de medicamento para verificar si existen resultados farmacogenómicos que contraindiquen la elección del medicamento o la dosis especificada.

Fuente: Clinical Decision Support (<https://www.healthit.gov/topic/safety/clinical-decision-support>) Fecha de revisión: 13 de Febrero de 2022

Los sistemas CDS pueden alertar a los clínicos acerca de información importante que surja con la variante del paciente previamente reportado, donde los proveedores de los sistemas de EHR no almacenan la secuenciación genética en estos registros, sólo vinculan la información con almacenes de datos genómicos (Aronson and Rehm 2015).

Los CDS pueden proporcionar a los médicos, el personal, los pacientes u otras personas conocimientos e información específica de la persona, filtrada de forma inteligente o presentada en los momentos apropiados, para mejorar la salud y la atención médica pues abarca una variedad de herramientas para mejorar la toma de decisiones en el flujo de trabajo clínico.

Estas herramientas incluyen alertas y recordatorios computarizados para proveedores de atención y pacientes; guías clínicas; informes y resúmenes de datos de pacientes enfocados; plantillas de documentación; soporte de diagnóstico e información de referencia contextualmente relevante, entre otras herramientas (“Clinical Decision Support” n.d.).

Los EHR y los sistemas usados por los médicos deberían tener acceso a los resultados de secuenciación genética directamente del laboratorio. Esto propicia que el diseño de interfases entre los EHR y los laboratorios se torne necesario, pues en la actualidad la mayoría de los resultados son transmitidos del laboratorio al proveedor por correo electrónico, lo cual dificulta la integración de los datos.

Actualmente, muchos grupos están trabajando en la interconectividad de los Sistemas CDS y las bases de datos de información genética. Uno de estos proyectos es DIGITize (*Displaying and Integrating Genetic Information Through the EHR Action Collaborative*), el cual convocó a partes interesadas de proveedores de gestión y

tecnología de la información sanitaria, centros de salud académicos, agencias gubernamentales y otras organizaciones para trabajar y examinar cómo la información genómica puede representarse e integrarse de manera uniforme en los registros médicos electrónicos en un formato estándar (Aronson and Rehm 2015).

El proyecto eMERGE tiene como objetivo principal el desarrollo, difusión y aplicación de enfoques que combinen los biorepositorios con expedientes médicos electrónicos para el descubrimiento genómico y la implementación de medicina genómica. Adicionalmente el proyecto incluye un enfoque social y ético, abordando temas como la privacidad, confidencialidad y las interacciones con la comunidad. Este proyecto actualmente se encuentra en fase III y consta de 9 sitios de estudio, dos instalaciones de secuenciación y genotipado y un centro de coordinación. Más específicamente, eMERGE Fase III tiene como objetivo:

- 1) Secuenciar y evaluar la implicación fenotípica de variantes raras en ~100 genes clínicamente relevantes que supuestamente afectan la función génica en aproximadamente 25 000 individuos
- 2) Evaluar las implicaciones fenotípicas de estas variantes,
- 3) Integrar variantes genéticas en EHR para atención clínica
- 4) Crear recursos comunitarios

Además, eMERGE III continuará evaluando el impacto en la salud, la rentabilidad y las implicaciones éticas, legales y sociales de informar variantes genéticas en una escala de población más amplia para pacientes, médicos e instituciones de atención médica (“Electronic Medical Records and Genomics (EMERGE) Network” n.d.).

Otra ventaja del uso de los EHR es que nos permite encontrar hallazgos fortuitos en su análisis a diferencia de un estudio prospectivo donde no es posible evaluar asociaciones entre variables aparentemente no vinculadas que no fueron establecidas desde el principio del estudio. Actualmente las mejores prácticas que utilizan los datos de los EHR están enfocadas en automatizar el proceso de análisis mediante el aprendizaje automático (*machine learning*). En un estudio (Becker *et al.*) se comparó la utilidad de 5 diferentes métodos de análisis: a) dos relacionados con la heurística clínica (que es la toma de decisiones clínicas basadas en escenarios anteriores que guardan mucha similitud con el ahora analizado), b) una con un modelo de regresión logística, (un tipo de análisis de **regresión** utilizado para predecir el resultado de una variable categórica en función de las variables independientes o predictoras) y c) dos de aprendizaje automatizado utilizando datos estructurados y no estructurados derivados de la EHR. Estos métodos de análisis mostraron mejores resultados frente a las estrategias tradicionales de fenotipificación (Glicksberg, Johnson, and Dudley 2018)

Por último, a pesar de las ventajas de los HRE respecto de otras metodologías de adquisición de datos, estos se limitan a la información obtenida durante la visita clínica, por lo que existen intervalos de tiempo entre las mediciones. Es por ello que, el uso de dispositivos electrónicos tales como relojes o teléfonos inteligentes pueden contribuir a recopilar la información de un paciente entre consultas medicas y contar con un mejor seguimiento de su padecimiento (Glicksberg, Johnson, and Dudley 2018).

8 Informática biomédica translacional

8.1 Infraestructura

La MP depende en gran medida de la disponibilidad de datos biomédicos fiables y de la capacidad de gestionar conjuntos de datos heterogéneos con altos niveles de dimensionalidad. La informática biomédica translacional (TBI, por sus siglas en inglés) es un campo emergente de la informática de la salud para avanzar en la medicina de precisión.

La creciente cantidad y variedad de información, por un lado, ha ampliado el arsenal para adquirir datos relevantes y, por otro lado, inevitablemente se ha sumado a la complejidad del análisis de estos datos. Esta riqueza de datos está haciendo que la comunidad de investigación, incluidos los científicos biomédicos y los médicos, estén bien informados sobre los antecedentes genéticos, genómicos, clínicos y ambientales del paciente.

La TBI abarca cuatro subdisciplinas: bioinformática, informática de imágenes, informática clínica e informática de salud pública(Chen, Lin, and Shen 2017).

La **bioinformática** está involucrada en procesos básicos desde el desarrollo de algoritmos para el análisis bioinformático de los datos de secuenciación de ADN, de datos de proteómica o metabolómica, las redes de interacción de diferentes moléculas, hasta la participación en el mantenimiento de las instalaciones donde almacena una gran cantidad de datos. Estos esfuerzos requieren de análisis gráficos avanzados y de vías combinando genotipo, expresión génica y otros datos, para

identificar vías desreguladas, inferir mecanismos y explicar la heterogeneidad de la enfermedad (Alzu'bi, Zhou, and Watzlaf 2014).

Otro de los aspectos de la TBI es la informática de imágenes, la cual utiliza enfoques de visualización para medir y monitorear la patogenia de enfermedades a nivel de tejido u órgano, centrado en la interpretación de la información derivada de los dispositivos de imagen.

Se espera que la informática de imágenes proporcione información médica de alto valor a partir de imágenes visuales para el diagnóstico y tratamiento tempranos de enfermedades. La gran cantidad de imágenes médicas también se han organizado y catalogado para el acceso público a través de los programas del NCI, como *Quantitative Imaging Network* y *The Cancer Imaging*.

Además de la recopilación de datos, estos archivos abiertos también implementan una serie de herramientas informáticas para permitir el intercambio de imágenes médicas entre múltiples grupos de investigación.

Por otra parte, la **bioinformática clínica** implica el uso de la informática en la gestión de datos clínicos personales. La informática clínica cubre una amplia gama de temas, incluida la evaluación de riesgos clínicos, el apoyo a las decisiones clínicas y la documentación clínica.

HR recopila la información clínica en papel de múltiples fuentes y la transforma a un formato digital de acuerdo con los estándares de interoperabilidad. EHR también proporciona sistemas de almacenamiento mediante los cuales los datos clínicos se organizan según categorías fenotípicas, un ejemplo de un programa de bioinformática clínica llamando INBIOMEDvision el cuál es recurso de coordinación para información

genómica, de imágenes y basada en información de la población(Chen, Lin, and Shen 2017).

El último elemento de la TBI es la **informática de salud pública** la cuál se centra en la información de salud basada en la población. La actividad de salud pública es extremadamente amplia, como la vigilancia de pandemias, el manejo de brotes, la prevención y la promoción de la salud, un ejemplo de este sistema es el Sistema de Vigilancia de Enfermedades (NEDSS, por sus siglas en inglés) el cuál esta facilita la transferencia eficiente de datos de salud pública a través de Internet. Promueve el intercambio oportuno de datos de vigilancia, informes de enfermedades y resultados de laboratorio entre funcionarios para guiar la toma de decisiones.

Es por ello, que el desarrollo de la bioinformatica necesita de analistas de datos moleculares altamente capacitados, bases de datos bien diseñadas, resultados de investigación precisos, algoritmos y programas de software y suficiente poder computacional para el análisis de datos masivos heterogéneos. En este momento, estos recursos normalmente no están disponibles para la mayoría de los médicos. Incluso si uno puede acceder fácilmente a excelentes programas de análisis de datos principalmente genómicos sigue siendo un desafío para la mayoría de los profesionales de la salud seleccionar el mejor programa y los parámetros correctos para ese conjunto de datos. En particular porque normalmente no cuentan con la formación necesaria en este campo. Por lo que será necesario formar especialistas que sean capaces de trabajar en grupos multidisciplinarios que tengan bases sólidas en biología general, anatomía y fisiopatología, programación de computadoras, diseño de bases de datos y estadísticas (Ahmed and Shabani 2019).

8.2 Bases de datos

8.3 Algoritmos y análisis de *Big data*

La investigación en la medicina de precisión genera una gran cantidad de datos heterogéneos a una gran velocidad debido al desarrollo de las tecnologías ómicas por lo que ya se considera fuente de *Big data*.

El término Big data biomédico se utiliza para designar todos los datos relevantes para la salud que pueden hacerse interoperables y, por lo tanto, susceptibles de extracción de datos predictivos con fines relacionados con la salud. La información puede ser generada por los servicios de salud, actividades de salud pública o investigación biomédica, inclusive se toma en cuenta datos como la exposición a factores ambientales tales como exposición solar o contaminación, información proveniente de aplicaciones fitness, redes sociales y dispositivos portátiles. Este tipo de información aporta un gran valor para realizar predicciones relacionadas con la salud tanto para individuos como para poblaciones (Vayena and Blasimme 2017).

El Big data biomédico se puede caracterizar por 5 propiedades: a) Voluminosa, por el número de datos que se generan en los análisis, b) Veloz, por la velocidad en la generación de la información, c) Veracidad como parámetro de la calidad de la información, d) Variedad de información disponible o tipos de datos determinado por la fuente de información y e) Valor intrínseco, definido como la habilidad de otorgar a los datos una importancia clínica (Kashyap et al. 2016)

Para desarrollar nuevo conocimiento científico proveniente de estudios de gran escala que explore fenotipos y endotipos, se deberá reunir toda esta información, analizar e integrar en múltiples niveles. Se estima que el volumen de datos de secuenciación genética ha crecido 10 veces cada año desde el 2002, como resultado de un

desarrollo en técnicas experimentales, dimensionalidad de la información y los análisis de alto rendimiento. Estos estudios ya no utilizarán formularios de informe de casos con 10 a 20 variables en 500 pacientes, medidos en un solo punto de tiempo. En cambio, cualquier variedad de datos de datos genómicos, transcriptómicos o de secuenciación podrían ser sumados a partir de una sola célula y aumentados por una variedad de factores como el número de órganos (O) ó sujetos (C) muestreados como se muestra en la Figura 9 (Seymour et al. 2017).

En lugar de 100 pacientes, los estudios de la medicina de precisión proponen el enrolamiento de más de un millón de pacientes en los que se toman muestras de múltiples órganos en más de un punto temporal. La escala de datos resultante, en la que buscar subgrupos importantes, superaría el billón (2×10^{12}) de puntos de datos. Ese aumento afectaría directamente el tiempo de cálculo para estimar incluso consultas simples. Los biólogos de sistemas computacionales deberán abordar esta gran cantidad de datos y utilizar software apropiado para coordinar e indexar conjuntos de datos ómicos (Seymour et al. 2017)

Figura 9.

Escala de datos posible mediante el muestreo de los datos generados a través de tecnologías ómicas.

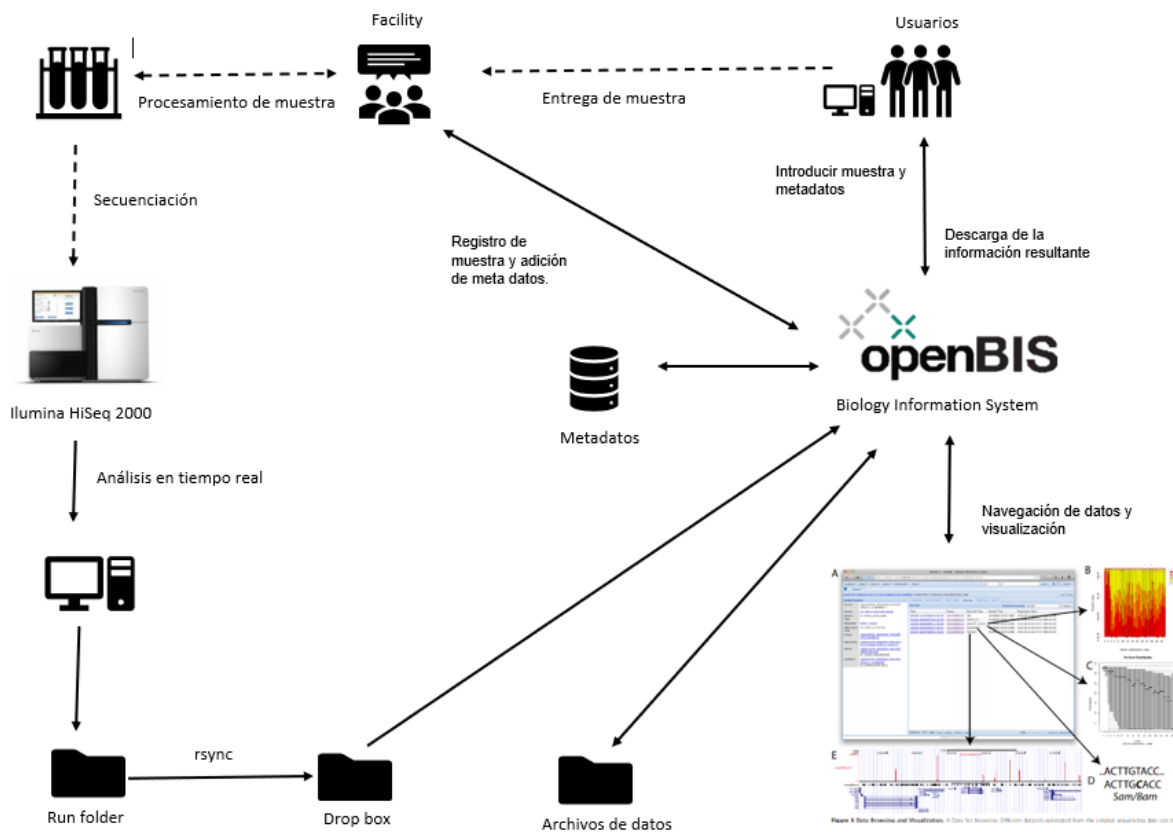
$$Escala\ de\ datos = \left[\sum_{i=1}^4 Di \right] \times O \times N \times T$$

<i>Donde,</i>	<i>Donde,</i>	<i>Donde,</i>	<i>Donde,</i>
D ₁ = Data del genoma	O ∈ {1, ...o}	N ∈ {1, ...n}	T ∈ {1, ...t}
D ₂ = Datos del transcriptoma	Si o es el	Si n es el	Si t es el
D ₃ = Datos del proteoma	número	número de	número de
D ₄ = Datos del metaboloma	organismos	pacientes	puntos
	muestreados	enrolados	muestreados

Fuente: Precision medicine for all? Challenges and opportunities for a precision medicine approach to critical illness por Christopher W. Seymour *et al.* (2017) Critical Care.

Para resolver el problema del gran volumen de información generada por las tecnologías se pueden utilizar sistemas de información dedicados específicamente al manejo de estos datos en gran escala. Por ejemplo, openBIS es un sistema de información que se puede utilizar para gestionar secuencias de ADN generadas por tecnologías NGS. Este sistema tiene dos componentes principales: un servidor de aplicaciones que se utiliza principalmente para gestionar metadatos (descripciones sistemáticas de datos), y el otro es un servidor de almacenamiento de datos de gran volumen, como secuencias del genoma con una interfaz que permite a los usuarios recuperar su información (Bauch et al. 2011).

Figura 10.
Esquema de sistema openBIS como herramienta para gestionar datos derivados de tecnologías NGS



Fuente: *openBIS: a flexible framework for managing and analyzing complex data in biology research* (p. 11) por Angela Bauch¹ et al., (2011), BMC Bioinformatics

La tarea de análisis se vuelve compleja pues se ha establecido que la mayor parte del desarrollo de las enfermedades no depende exclusivamente de los cambios en un gen. La causa de la enfermedad puede implicar: a) cambios en múltiples genes en múltiples cromosomas, desde las regiones codificantes de proteínas de un gen hasta regiones intragénicas (intrones), elementos reguladores, regiones intergénicas, b) genes que codifican proteínas hasta otros tipos de ARN (por ejemplo, ARNt, microARN), c) aspectos epigenéticos de genes vinculados con la enfermedad, d) polimorfismos de un solo nucleótido (SNP, o mutaciones puntuales) hasta inserciones y deleciones cortas, variaciones en el número de copias y otras variaciones estructurales, y e) genes individuales hasta interacciones gen-gen e interacciones gen-ambiente. Por otra parte, con el desarrollo de técnicas biotecnológicas, instalaciones computacionales más poderosas y algoritmos de análisis e integración de datos más robustos, los investigadores descubrirán otros factores involucrados en el desarrollo de enfermedades diferentes a los mencionados, y se sumarán a la comprensión actual del origen de las enfermedades.

Hay que considerar que la medicina de precisión deberá procesar los resultados útiles de su investigación de forma organizada a los médicos para que les sea útil en el diagnóstico, pronóstico y tratamiento de una enfermedad. Es poco viable solicitar a los profesionales de la salud que busquen todos los artículos relacionados con la enfermedad, sobre todo cuando el resultado final va acompañado de características muy específicas para cada paciente.

8.4 Seguridad informática

Los datos moleculares de los pacientes son altamente sensibles y necesitan ser protegidos pues no solo contienen información acerca de un paciente en particular. Al

incluir datos obtenidos de las técnicas de secuenciación también se incluye información potencial de los ascendientes y descendientes por muchas generaciones de la familia del paciente. Se ha demostrado que se puede utilizar un algoritmo inteligente y datos recopilados de bases de datos públicas y redes sociales, para la identificación de forma única del propietario de una secuencia de genoma anónima almacenada en bases de datos públicas (Alzu'bi, Zhou, and Watzlaf 2014). Los profesionales involucrados en el manejo de información de la salud (HIM, *Health information manager*, por sus siglas en inglés) juegan un rol muy importante en la medicina de precisión porque son los responsables de manejar y proteger los datos biometricos de los pacientes para extraer información útil y respetar el derecho a la privacidad de todo paciente. Para ello deben tener ciertas habilidades y conocimientos en el manejo de la información molecular, conocimiento del manejo de una base datos, encriptamiento y desencriptamiento de datos sensibles, análisis estadístico, regulaciones del uso de datos de pacientes y establecer políticas y procedimientos para el manejo de problemas legales y éticos relacionados con los registros de paciente (Alzu'bi, Zhou, and Watzlaf 2014)

Para reforzar la integridad y seguridad de los datos y registros médicos, algunas empresas proponen implementar la tecnología *blockchain*, la cual es una tecnología mediante una base de datos descentralizada y digitalizada en un libro mayor público de todas las transacciones en la red. La característica clave es una cadena de bloques, que es la base de datos distribuida. Esta base de datos cuenta con muchas copias en varios sistemas informáticos, lo que crea una red de igual a igual que indica que ya no hay un organismo centralizado que controle el acceso a los datos (Ahmed and Shabani 2019).

Aunque los sistemas y empresas han establecido medidas físicas, organizacionales y técnicas para garantizar la seguridad de la información durante el proceso de recolección, transmisión y una vez recibida, debemos considerar que ninguna medida de seguridad es 100 % segura por lo cual no pueden garantizar la seguridad de la información (Ahmed and Shabani 2019).

Por lo tanto, es posible que se necesiten medidas de seguridad más robustas para la protección de datos genómicos personales, y este sistema debe configurarse antes de la amplia aplicación de la información genómica en la práctica clínica. Por lo que, será necesario desarrollar medidas de seguridad robustas y sofisticadas que cubran la necesidad de la protección de los datos personales en las bases de datos sin que se impida el acceso a los investigadores que le pueden otorgar una perspectiva clínica de utilidad (Alzu'bi, Zhou, and Watzlaf 2014).

9 Legislación

9.1 Normas, leyes y tratados

Los recursos referentes a muestras biológicas, humanas o datos de salud se han convertido en bienes muy valiosos para la investigación referente a la salud. De manera que se incentiva la donación de muestras y el acceso a estos recursos, por ello se han realizado una serie de cambios regulatorios en el proceso de la obtención de consentimientos, los cuales incrementan el grado de control que ejercen los sujetos a quien pertenece los datos, fomentando que se sientan seguros de compartir o revelar información para los fines que a ellos convengan.

Debido a que, la medicina de precisión necesita un amplio acceso la *información de salud personal* que se define como todo aquel dato relativo al estado de salud de un individuo identificable y que incluye los datos biométricos del individuo, expediente

clínico electrónico, datos de estilo de vida, así como todos los datos disponibles que provengan de tecnologías ómicas, etc., esta debe ser regulada (Vayena and Blasimme 2017).

Actualmente, las discusiones acerca del consentimiento y el uso de los datos de los pacientes se limitan principalmente a los derechos individuales en la donación de información y no abordan suficientemente los intereses pertinentes de los miembros de la familia. A nivel práctico, también es crucial investigar hasta qué punto los miembros de la familia deben participar en el proceso de intercambio de datos genómicos personales, incluido el consentimiento informado. Actualmente, los marcos legales aplicables para la protección de datos personales se limitan predominantemente a reconocer a las personas como “sujetos de datos”, es decir, son propietarios independientes de sus datos y determinan cómo quieren que sean usados, no obstante, esto no se extiende a los miembros de la familia (Ahmed and Shabani 2019).

Por lo que: la medicina está convirtiéndose en una enorme fuente de datos personales, en donde el consentimiento informado y la confidencialidad médica ofrecen mecanismos limitados de control sobre la producción, recopilación, uso y circulación de la información de los pacientes. Es conveniente explorar otras alternativas que aumenten el control individual del surgimiento de los derechos de portabilidad de datos, como un mecanismo para controlar el acceso a los datos, así como nuevos procedimientos de consentimiento informado, como herramientas para controlar el uso de datos (Vayena and Blasimme 2017).

El manejo de los datos puede describirse con tres conceptos fundamentales: acceso, control y gobernanza (figura 11). En el caso del primero se hace alusión a la habilidad de filtrar el acceso a los datos de salud o información personal por organizaciones o personas no autorizadas. En el caso del segundo, aparte de definir quien tiene acceso a esta información, también incluye la descripción de para que serán utilizados estos datos, y si están alineados con los intereses y expectativas de quien los proporcionó. Por último, la gobernanza hace referencia a la construcción de las estructuras entre las instituciones, el estado y los pacientes, de manera que se pueda proporcionar confianza (Vayena and Blasimme 2017).

El acceso a la información del individuo conlleva riesgos éticos, sociales y legales, por lo que es importante que se consideren las regulaciones existentes sobre los datos del paciente, los procedimientos y políticas para manejar estos problemas, para implementar la medicina de precisión respetando los derechos de los pacientes

Figura 11

Tres dimensiones del control de datos.



Fuente: *Biomedical Big Data: New Models of Control Over Access, Use and Governance* (p. 501-513) por Effy Vayena y Alessandro Blasimme (2017), *Bioethical Inquiry*.

La legislación de datos personales define y promueve los derechos e intereses de los sujetos o individuos con respecto a su información personal. En el campo del cuidado de la salud, Estados Unidos cuenta con la *Health Insurance Portability and Accountability Act (HIPAA)*, promulgada desde 1996 y que tiene como objetivo definir las restricciones de los actores involucrados en el procesamiento de los datos de salud. Por su parte en Europa existe una directiva similar desde 1995 y que fue remplazada por la actual: *General Data Protection Regulation (GDPR)* en 2016, en donde se plantea el derecho a la portabilidad, la cual le permite acceder los datos referentes a él o ella, por ejemplo, solicitar sus archivos clínicos electrónicos o datos recolectados y obtenidos mediante dispositivos médicos. Este derecho no sólo se limita al acceso del sujeto a su información y la rectificación de la misma, sino también para convertirla en disponible de una institución de investigación a otra.

Respecto a las restricciones de uso de los datos de salud, existe en Estados Unidos una guía elaborada por el *U.S Department of Health and Human Services (DHHS)*, por sus siglas en inglés) y *The Food and Drug Administration (FDA)*, por sus siglas en inglés) para el uso de consentimiento informado electrónico. Esta guía presenta una ventaja, pues su formato electrónico permite mejorar la divulgación de la información referente a las condiciones de uso de los datos personales. Ya que la mayor parte de la captura y procesamiento de esta información es mediante plataformas online, resultaba necesario el desarrollo de un mecanismo de consentimiento electrónico.

Un ejemplo de esto es el presentado por la compañía americana *Sage Bionetworks*, la cual es una organización de investigación sin fines de lucro que busca acelerar la traducción de la ciencia en medicina y que desarrollo el eConsent, el cual tiene como objetivo maximizar la comprensión e información a los voluntarios en estudios de

investigación a través del uso de recursos visuales como iconos y animaciones, particularmente está enfocado en la información se obtiene mediante aplicaciones o Apps. Este desarrollo conlleva la participación de ingenieros, desarrolladores de software, médicos, investigadores, pacientes, representantes de pacientes, tomando en cuenta a su vez la experiencia de uso con respecto a los “Términos de uso”, “Políticas de privacidad” y otros consentimientos para aplicaciones de teléfonos inteligentes (Doerr, Suver, and Wilbanks 2016).

Lo anterior generó como resultado la identificación de las características obligatorias para estar en cumplimiento con los marcos regulatorios, además de otras características deseables que mejoran la transparencia y consistencia de los datos. Dicho consentimiento fue asesorado y aprobado por asesores en bioética, investigadores, pacientes y miembros de los *Institutional Review Boards* (Doerr, Megan and Suver, Christine and Wilbanks, John. 2016)

México también cuenta con regulaciones referentes a la privacidad de datos y derechos de los pacientes entre otras cuestiones, que pueden servir de base para implementar la medicina de precisión y que se van a mencionar a continuación

- 1) **Ley General de Salud (LGS). Título quinto: Investigación para la salud.** Este título realiza una descripción de las actividades o acciones que están encaminadas al desarrollo de la investigación científica y tecnológica en materia de salud, estableciendo en su artículo 100 aspectos fundamentales para la protección de los participantes o pacientes en las actividades de investigación. Considera que sólo podrá realizarse la investigación si el conocimiento a producirse no puede obtenerse mediante otra metodología, así como que sólo podrá efectuarse cuando

exista una razonable seguridad que no expone a riesgos ni daños innecesarios al sujeto en experimentación. Menciona que se necesita la firma del consentimiento por escrito del sujeto en quien se realizará la investigación, o de su representante legal, y precisa que este tipo de investigaciones sólo podrán realizarse por profesionales de la salud en instituciones médicas que actúen bajo la vigilancia de las autoridades sanitarias competentes, entre otros aspectos.

2) Reglamento de la Ley General de Salud en materia de Investigación para la salud. Título Segundo. De los aspectos éticos de la Investigación en seres humanos. Se plantean las disposiciones comunes relativas a la investigación en la que el ser humano sea sujeto de estudio, tomando en consideración el respeto a su dignidad y la protección de sus derechos y bienestar. Señala que la Investigación que se realice en seres humanos deberá desarrollarse conforme a las siguientes bases:

I. Deberá adaptarse a los principios científicos y éticos que justifican la investigación médica, especialmente en lo que se refiere a su posible contribución a la solución de problemas de salud y al desarrollo de nuevos campos de la ciencia médica; Fracción reformada DOF 02-04-2014; II. Se fundamentará en la experimentación previa realizada en animales, en laboratorios o en otros hechos científicos; III. Se deberá realizar sólo cuando el conocimiento que se pretenda producir no pueda obtenerse por otro medio idóneo; IV. Deberán prevalecer siempre las probabilidades de los beneficiarios esperados sobre los riesgos predecibles; V. Contará con el consentimiento informado del sujeto en quien se realizará la investigación, o de su representante legal, en caso de incapacidad legal de aquél, en términos de lo dispuesto por este Reglamento y demás disposiciones jurídicas aplicables; Fracción reformada DOF 02-04-2014; VI.

Deberá ser realizada por profesionales de la salud a que se refiere el artículo 114 de este Reglamento, con conocimiento y experiencia para cuidar la integridad del ser humano, bajo la responsabilidad de una institución de atención a la salud que actúe bajo la supervisión de las autoridades sanitarias competentes y que cuente con los recursos.

Otra de las regulaciones vigentes y con mayor relevancia para esta materia es la que se refiere al acceso y protección de datos personales y datos personales sensibles. Los datos personales están definidos como cualquier información concerniente a una persona física, que la identifique o la haga identificable y los datos personales sensibles como aquellos datos que afecten a la esfera más íntima del titular y cuya utilización indebida pueda dar origen a la discriminación o conlleve un riesgo grave. Por ejemplo, los datos de origen étnico, estado de salud (pasado, presente o futuro), información genética, creencias religiosas, opiniones políticas o preferencia sexual. En México la ley destinada a regular el acceso y manejo a esta información está definida en la Ley Federal para la Protección de Datos Personales en Posesión de Particulares (LFPDPPP), así como en la ley General de Protección de Datos Personales en Posesión de Sujetos Obligados. La primera ley es aplicable para personas físicas o morales, y la segunda a cualquier autoridad, entidad, órgano y organismo de los Poderes Ejecutivo, Legislativo y Judicial, órganos autónomos, partidos políticos, fideicomisos y fondos públicos.

Con la entrada en vigor de las leyes antes mencionadas se persigue que garanticen un manejo adecuado de los datos personales en favor de la privacidad y de la autodeterminación informativa de los titulares de esta información. De acuerdo con éstas, los principales ejes sobre la protección de datos personales son: Licitud,

Lealtad, Información, Consentimiento, Finalidad, Proporcionalidad, Calidad y Responsabilidad. Estos principios son el fundamento de las obligaciones que tienen los responsables del tratamiento de los datos personales en el ejercicio de sus actividades, de la misma manera debe quedar garantizado el ejercicio de los titulares de los datos respecto a sus derechos ARCO (que son las siglas de acceso, rectificación, cancelación y oposición). Además de los derechos ARCO, la LDPDPPP otorga a los titulares el derecho a revocar el consentimiento otorgado previamente para el tratamiento de su información personal, en cualquier fase del tratamiento.

Otro principio de especial importancia es el de consentimiento, en el cual se debe contar con el permiso del titular para el tratamiento de sus datos personales. Este consentimiento puede ser tácito, expreso o expreso por escrito dependiente del tipo de datos. El primero es aplicable para datos personales, el segundo para datos financieros, es decir no aplicable para el tema de este trabajo y el tercero para datos personales sensibles.

El principio de información es de gran importancia pues en él se establece la obligación del responsable a comunicar al titular de los datos personales, las características del tratamiento al que será sometida su información personal, así como los medios para ejercer sus derechos concretado a través del *Aviso de confidencialidad*. Este aviso debe redactarse conforme a los elementos que señala la norma aplicable, ya sea la LFPDPPP o la LGPDPPSO. Su redacción debe ser de manera sencilla, con la información necesaria, utilizando un lenguaje claro y con una estructura que facilite su comprensión.

Estas mismas legislaciones hacen mención de la *Evaluación de Impacto* en la protección de datos personales, el cual es un análisis documentado mediante el cual

los responsables de tratar los datos colocan y analizan la operación del sistema con base en las políticas públicas, programas, sistemas o plataformas informáticas, y valoran los impactos reales respecto de determinado tratamiento de datos, con el objetivo de identificar posibles riesgos para el uso de esta información e implementar medidas de protección que mitiguen los riesgos identificados.

Con base en lo anterior, es necesario enfatizar que las protecciones legales para los datos en conjunto con los innovadores enfoques tecnológicos deben garantizar sistemas de información seguros como desidentificar y difuminar datos de salud, poniendo como una opción el controlar el acceso a través de *blockchain*, mencionada anteriormente. De esta forma los registros no pueden modificarse retroactivamente sin el consenso de toda la red implicada. Otro filtro de seguridad puede relacionarse con la vinculación de datos usando identificadores *hash*, los cuales son algoritmos que pueden transformar cualquier dato en una nueva serie de caracteres de longitud fija que preservan la privacidad, así como el análisis de datos usando cifrado homomórfico, en donde se puede realizar la manipulación de los datos sin tener que romper el cifrado de los mismos (Denny and Collins 2021).

En conclusión, el uso de la información personal de salud demanda que las instituciones de investigación identifiquen claramente y declaren quien es responsable del manejo de los datos, garanticen el almacenamiento seguro y establezcan los procedimientos de acceso, así como realicen el monitoreo del uso de esa información. Estableciendo lo anterior, los individuos se sentirán motivados para que su información personal este disponible bajo un marco legal confiable que haga posible el análisis de sus datos aplicables al mejoramiento del cuidado de la salud (Vayena and Blasimme 2017).

9.2 Bioética

La bioética es una disciplina que se ha consolidado como un espacio de reflexión multidisciplinario y necesario para abordar con éxito los problemas complejos suscitados en el ámbito de la atención sanitaria y la visión humanista centrada en la ética, apoyándose en otras disciplinas, por ejemplo: biológicas, humanistas, económicas, filosóficas, políticas y legislativas con el objetivo de buscar el bien social.

En la actualidad uno de los mayores retos a considerar es la falta de diversidad de poblaciones involucradas en los estudios clínicos. Menos del 3 % de los participantes de estudios de asociación del genoma completo corresponden a poblaciones africanas, hispanas o latinoamericanas, mientras que 86 % corresponden a poblaciones caucásicas. Con la creciente profundidad de los datos, tenemos la oportunidad de reemplazar los rasgos de raza y etnicidad por medidas más específicas, particularmente el concepto de “raza” conlleva aspectos sociales, culturales, políticos, geográficos y biológicos que podrían perpetuar el racismo. Sin embargo, si se realiza una recolección de los determinantes sociales en salud en la investigación y el cuidado clínico, en combinación con medidas más específicas acerca de la influencia ambiental, hábitos y genética ancestral se obtendrán mejores datos y orientaciones muestrales para la estratificación de riesgos y tratamientos adecuados.

Otras cuestiones legales, sociales y éticas están relacionadas con la aplicación de la información personal en la medicina. Por ejemplo, algunas personas pueden verse afectadas al no poder obtener un seguro médico porque los resultados de sus pruebas genómicas no se mantienen confidenciales.

Como ya se ha comentado, esta información puede ser de gran utilidad para la investigación, en el desarrollo de nuevos tratamientos y mejoramiento de los diagnósticos. Aunque se ha observado que no siempre los pacientes están dispuestos a compartirla, en una encuesta se determinó que entre los participantes existe una disposición positiva para compartirla con organizaciones dedicadas a la investigación, sobre todo públicas. Con las de financiamiento privado con fines de lucro, aún existe cierta resistencia, y definitivamente no están dispuestos a compartirla con aseguradoras y empresas de publicidad. Por lo cual se debe otorgar al individuo la certeza de tener el control del uso de su información, a manera que se garantice un estado de acceso controlado y se evite la sensación de riesgo de intrusión de la privacidad. (Ahmed and Shabani 2019)

Muchos de los clientes de empresas de secuenciación del genoma, hoy denominadas DTC (*Direct-To-Consumer*, por sus siglas en inglés) como 23andMe han accedido a compartir esta información con fines de investigación, aunque al aceptar su participación renuncian a los derechos sobre los productos o beneficios derivados de la investigación, aunque provengan de su información. Esto es percibido como injusto por parte del paciente (Ahmed and Shabani 2019)

Por lo anterior y ante la creciente necesidad del acceso a estos datos como recursos para la investigación, se ha creado el denominado “*DNA data marketplace*”, donde empresas como Nebula Genomics, LunaDNA o EncrypGen, otorgan a las personas incentivos, los cuales pueden ser monetarios, para secuenciar su genoma y compartir la información con las partes comerciales interesadas (Ahmed and Shabani 2019)

En el caso de Nebula Genomics, su modelo establece que los clientes adquieran y almacenen su propia información genómica y en caso de compartirla adquieran

compensaciones. Nebula Genomics actúa como facilitador para compañías farmacéuticas en el reclutamiento de participantes con características de interés para la investigación. Esto lo realiza mediante la búsqueda anónima de pacientes, que cuando son contactados por las compañías farmacéuticas pueden decidir si comparten su información genómica y otros datos médicos. Otras empresas, como EncrypGen, declaran que su objetivo es romper la barrera de la retención de genómicos de los clientes de las empresas de DTC y las empresas de investigación y desarrollo mediante el empoderamiento de los usuarios con su información. Para ello le brindan la opción de monetizarlo al compartirlo con terceros interesados. Por su parte, LunaDNA ofrece acciones de su compañía al usuario al contribuir con sus datos genómicos, así como por cargar sus informes médicos y estilo de vida. Esas acciones otorgan participación en las ganancias de la investigación y el desarrollo médicos. Los incentivos están en función de la cantidad de información compartida.

Es aquí donde nos enfrentamos a diversas problemáticas éticas como la rescisión del consentimiento del uso de los datos del paciente. Esta acción de rescisión debería poder realizarse en cualquier momento de la investigación y siempre que el paciente lo solicite. Sin embargo, surgen preguntas acerca del retiro de la información, cuando el acceso estuvo motivado por incentivos que ya fueron entregados al beneficiario. Por lo anterior, empresas como LunaDNA establecen que se requiere consentimiento para el uso continuo de la información para que no finalice el ser propietario de las acciones que se le hayan emitido en relación con la contribución de esos datos compartidos.

Una forma de resolver y establecer los derechos de los pacientes o participantes consiste en conocer cuál es el punto de vista o expectativas de los participantes en la

investigación como vía para dar reconocimiento a su participación. Esto podría implicar dar a conocer de manera obligatoria los resultados de la investigación o informarle al participante si existen hallazgos incidentales posteriores. Esto puede otorgar a los sujetos un papel más activo, logrando un estímulo para que las personas contribuyan al desarrollo de la medicina de precisión donando muestras y aceptando la recopilación de datos sobre sí mismos (Blasimme and Vayena 2016).

Una de las regulaciones más importantes que explica los principios éticos fundamentales para la participación de humanos en la investigación es el Informe Belmont, el cual fue creado por el Departamento de Salud y Bienestar de los Estados Unidos titulado "Principios éticos y pautas para la protección de los seres humanos en la investigación", acuñando su actual nombre en 1979. Actualmente es una referencia para los investigadores para asegurar que se cumplan con las regulaciones éticas, siendo generalmente tarea del Comité de ética e investigación el cumplimiento cabal de los lineamientos establecidos en esta declaración. Algunos de los derechos fundamentales descritos en este documento son:

Respeto a las personas, protegiendo la capacidad que tienen de decidir con libertad si desean participar en el estudio o no, una vez que les han sido explicados todos los riesgos, beneficios y potenciales complicaciones. En el contexto de la salud, estos mecanismos son considerados la piedra angular para la construcción y el mantenimiento de la confianza en los sistemas de salud y los institutos de investigación (Vayena and Blasimme 2017).

Este principio conlleva la obtención de un consentimiento informado donde se debe otorgar una amplia explicación de la investigación, así como el establecimiento del derecho de retirarse del estudio cuando lo desee. Beneficencia, que implica la

búsqueda del incremento de los beneficios para los sujetos y la disminución de los riesgos. Justicia, donde los riesgos y beneficios de un estudio de investigación deben ser repartidos equitativamente entre los sujetos de estudio (Zucker 2007).

Dicho documento considera tres importantes elementos: a) Información, para asegurar que los sujetos reciben la información suficiente incluyendo: procedimiento de investigación, propósito, riesgo y beneficios que se han previsto, si existen procedimientos alternos y una declaración que ofrece al sujeto la oportunidad de hacer preguntas, así como la oportunidad de retirarse de la investigación si así lo desea; b) Comprensión, se debe adaptar la presentación de la información a las capacidades del sujeto, siendo responsabilidad de los investigadores averiguar si el participante ha entendido la información realizando incluso una confirmación por medio de pruebas orales o escritas, en el caso en que la comprensión está severamente limitada como ocurre en casos de inmadurez o incapacidades mentales, debe tomarse alguna disposición especial, pero aun así se debe respetar el derecho a decidir; c) Voluntariedad, donde se establece que el consentimiento de participar cobra validez sólo si es otorgado voluntariamente para lo cual debe estar libre de situaciones de coerción (Zucker 2007).

Otra recomendación importante es evaluar en qué circunstancias se ofrecen incentivos financieros a cambio de datos personales relacionados con la salud. Esto puede amenazar la validez del consentimiento y comprometer la capacidad del paciente a responder con la verdad, así como hacer un estudio de cómo los participantes sanos con diversos contextos socioeconómicos responden ante el estímulo financiero a cambio del intercambio de datos. En el modelo tradicional de investigación, el comité de ética es quién evalúa los riesgos de ofrecer estos

incentivos, sin embargo, en el mercado del DNA no está presente esta evaluación (Ahmed and Shabani 2019).

Es importante también establecer las limitaciones de los consentimientos informados, pues los participantes que voluntariamente otorgan muestras y datos para la construcción de biobancos no conocen a futuro quien tendrán la posibilidad de acceder a esta información y con que propósitos, ya que los términos en los que acepta se presentan con gran detalle o se presentan con ambigüedad (Vayena and Blasimme 2017).

La manera correcta sería empoderar a los pacientes e individuos informándose adecuadamente acerca de las limitaciones en el control de la información una vez que ya fue compartida con un tercero.

10 Beneficios de la implementación

10.1 Beneficios médicos

Desarrollo de nuevos tratamientos

Estas incluyen grandes estudios de cohorte para el trabajo de descubrimiento de fenotipos y la identificación de objetivos candidatos mediante el modelado de sistemas integrados. El seguimiento incluiría la traducción de fenotipos clínicos y biológicos en mecanismos biológicos putativos, la modelización de estos endotipos en modelos animales estrechamente vinculados a los humanos y estudios preclínicos in vitro de terapias candidatas (Seymour et al. 2017). Un ejemplo es el trabajo preclínico cuyo objetivo fue identificar posibles dianas farmacológicas para el tratamiento de la sepsis. Estudios clínicos anteriores sugieren que existen subgrupos de pacientes con un control adecuado de la infección que desarrollan fenotipos

inflamatorios únicos como: insuficiencia orgánica múltiple asociada a trombocitopenia (*TAMOF*, por sus siglas en inglés), disfunción orgánica múltiple secuencial (*SMOF*, por sus siglas en inglés), y el síndrome de activación de macrófagos o inmuno parálisis (*MAS*, por sus siglas en inglés). Estos fenotipos tienen una patobiología distinta, factores de riesgo específicos y respuestas a biomarcadores a la terapia in vivo. Un paso más fue un análisis secundario de un ensayo fase III del antagonista del receptor IL-1 (Anakinra). En pacientes sépticos con criterios similares al síndrome de activación de macrófagos (es decir, coagulación intravascular diseminada y / o disfunción hepatobiliar), se observó una respuesta favorable al bloqueo de IL-1ra en la enfermedad reumatológica, y hubo una reducción absoluta de 30 % en la mortalidad del tratamiento frente al placebo. Esto no se observó en pacientes sin características de MAS.

El desarrollo de un nuevo medicamento también podría enfocarse desde etapas tempranas en pacientes que sean más beneficiados, disminuyendo los efectos adversos y reduciendo los costos, estas diferencias podrían ayudar con una rápida aprobación regulatoria y la pronta adopción de los médicos de los nuevos tratamientos (National Research Council et al. 2012).

Las similitudes moleculares entre enfermedades aparentemente no relacionadas serían de relevancia directa en el descubrimiento de nuevos medicamentos, al dar lugar a una investigación específica acerca de rutas relevantes para la enfermedad que son compartidas entre enfermedades relacionadas molecularmente (Seymour et al. 2017)).

La identificación de relaciones entre los factores ambientales y las decisiones de estilo de vida con la incidencia de la enfermedad pueden hacer posible la reducción de la incidencia con intervenciones en el estilo de vida.

- **Detección de una enfermedad y diagnóstico.**

Si tenemos una red de conocimiento en la cual las enfermedades están definidas en términos de vías moleculares, nos proporciona un potencial para acelerar el descubrimiento de los mecanismos subyacentes de las patologías. Las similitudes entre los perfiles moleculares de enfermedades con un mecanismo patogénico conocido y otro mecanismo patogénico desconocido pueden apuntar hacia el esclarecimiento de mecanismos compartidos, o constituir una guía para resolver la interrogante acerca de vías celulares que involucradas en la patogénesis de ambas enfermedades (National Research Council et al. 2012).

- **Predisposición a una enfermedad.**

El seguimiento de los resultados en pacientes individuales a lo largo del tiempo permitirá comprobar y, en el mejor de los casos, verificar el valor pronóstico de las clasificaciones de base molecular. La utilidad clínica de identificar las predisposiciones depende de la disponibilidad de intervenciones que prevengan, retrasen el inicio de la enfermedad o quizás mejoren la severidad de la enfermedad.

10.2 Beneficios económicos

La medicina de precisión se ha empezado a convertir en algunos países en una práctica de rutina, sin embargo, el aspecto costo-efectividad es uno de los obstáculos, sin que ello signifique que no pueda ser una práctica rentable en los próximos años. Para evaluar el valor de la medicina de precisión deben hacerse dos consideraciones

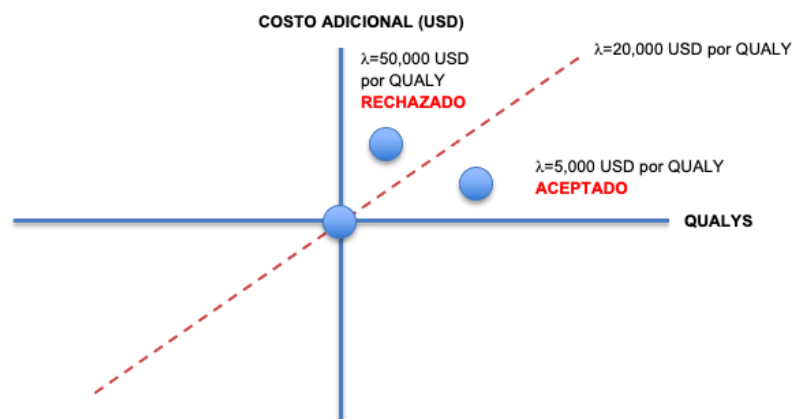
importantes: el costo inicial y el valor terapéutico que aportan para el paciente por el resto de su vida. Estos aspectos están mediados por aspectos éticos y sociales. Entonces el desarrollo de un modelo económico para justificar el costo de la medicina de precisión requiere de un nuevo planteamiento. Este constituiría un nuevo modelo de negocio por parte de las compañías farmacéuticas, los gobiernos y los pacientes. Dependiendo de la divisibilidad (costo del medicamento separado del costo del tratamiento general) así como el valor de horizonte de generación de valor (tiempo en que se amortiza el costo) son los elementos que deberán incluir y desarrollar los modelos de negocio para romper con las limitantes de presupuesto de corto plazo. El reto de la medicina de precisión es entonces convencer a los patrocinadores que esta práctica es viable y puede ser financiada mediante la amortización a largo plazo, la distribución de costo entre los actores implicados y los incentivos que pueden ser regidos por indicadores de resultados.

Uno de los modelos más aceptados para realizar un análisis de costo-efectividad está basado en la comparación de los costos y los resultados del sistema de salud, el cual es medido mediante QALY (*Quality Adjusted Life Years*, por sus siglas en inglés), la cual es una métrica compuesta de indicadores de salud, calidad y duración de vida, que permite comparar intervenciones en enfermedad completamente diferentes, para evaluar al sistema de salud a lo largo del tiempo. Entonces, cuando se introduce una nueva tecnología/tratamiento, se calcula y compara la relación entre el costo incremental de la nueva tecnología y el QALY incremental (tasa costo-efectividad incremental) o ICER, (*Incremental Cost-effectiveness Ratio*, por sus siglas en inglés) con un umbral predefinido (Figura 12). Un tratamiento se pone a disposición o reembolsa sólo si el ICER está por debajo del umbral, garantizando que un nuevo

tratamiento no excluye tratamientos que tengan un mayor valor social (“The Economics of Precision Medicine” n.d.)

Figura 12.

El plano de rentabilidad ilustra los costos incrementales (eje Y) y las consecuencias incrementales (eje X) entre una nueva tecnología sanitaria (por ejemplo, medicina de precisión) y un comparador relevante (por ejemplo, práctica actual). La línea discontinua que atraviesa el origen del plano representa el umbral de rentabilidad.



Fuente: *The economic case for precision medicine* (Vol. 3 p. 1-9), por P. Galvan Sean et al., (2018), Expert review of precision medicine and drug development.

Teniendo esto en cuenta la evaluación mediante ICER no es sencilla, para la medicina de precisión, pues es un método relativamente sensible al estándar de tratamiento. Para el caso de las enfermedades cardiovasculares, el estándar en medicamentos son las estatinas e inhibidores de la ECA (enzima convertidora de angiotensina) los cuales son baratos pues ya cuentan con múltiples opciones genéricas. Los nuevos medicamentos como Entresto, presentan tasas desfavorables de ICER por los bajos costos del estándar de cuidado, orillando la innovación hacia indicaciones donde los precios son altos.

El ICER no tiene consideraciones equitativas. A pacientes con tratamientos costosos se niega el acceso al mismo en caso de que el resultado no sea suficientemente bueno desde la perspectiva ICER. Como los QALYs son conducidos principalmente por la esperanza de vida ganada, los tratamientos para enfermedades con bajas tasas de mortalidad como epilepsia o psoriasis presentan tasas pobres de ICER, limitando el acceso a tratamientos innovadores.

La medicina de precisión también ha planteado importantes preocupaciones económicas cuando se debe incidir en su inversión, pues aún en naciones con un elevado índice de PIB, las pobres tasas con las metodologías de evaluación costo-beneficio de las terapias no son prometedoras, pues algunos tratamientos solo ofrecen prolongar la supervivencia unos meses contra el tratamiento convencional.

A continuación, se abordan algunos de los costos asociados a la medicina de precisión

- 1) Enrolamiento y selección de pacientes
- 2) Estratificación (incluida la secuenciación genética) que incluye la recolección, análisis, almacenamiento y distribución de la información.
- 3) Intervención, tratamiento, adaptación de medicamentos existentes que incluye nuevos tratamientos (Cambios de dosificación, formulación, protocolos), nuevos medicamentos dirigidos a individuos con biomarcadores identificados y nuevos medicamentos para aquellos que no tienen respuesta a los existentes.
- 4) Distribución de la información médica actualizada.
- 5) Adherencia del paciente y monitoreo del tratamiento (“The Economics of Precision Medicine” n.d.)

Una vez planteado los aspectos generales de los costos de implementación, contrastaremos con las ganancias económicas que esta práctica puede representar:

1. *Realizar y mejorar el seguimiento de los costos reales del cuidado de la salud*, de acuerdo con la OMS del 20 % al 40 % del gasto se desperdicia dando tratamiento a población que no va a responder. La MP puede eliminar este desperdicio mediante terapias dirigidas a población a la cual si le será de beneficio el tratamiento asignado.
2. *Valor de los resultados*, la inversión inicial de una prueba de diagnóstico temprana y precisa puede ahorrar dinero a largo plazo mientras incrementa la calidad de vida de los pacientes (aspecto no evaluado por el ICER).
3. *La prescripción de medicamentos basados en pruebas genéticas de proteínas metabolizantes*, permitiendo reducir los eventos adversos sobre todo en pacientes poli tratados, generalmente pacientes ancianos.
4. *Estudios clínicos*, se pueden volver analizar los datos de estudios clínicos fallidos, para generar nuevas perspectivas utilizando como herramientas el nuevo conocimiento generado y la correlación con otras bases de datos.

Al plantear un enfoque basado en el valor, tenemos dos perspectivas a considerar, la primera se basa en considerar el valor del conocimiento y la información generada como herramienta para disminuir la incertidumbre gracias a los diagnósticos precisos, el valor de una esperanza de cura, una opción de prolongación de la vida abre una opción al paciente de un beneficio por innovaciones futuras.

Por otra parte, desde un aspecto fiscal existe el enfoque del impacto macroeconómico de la enfermedad y salud de una población. Desde esa visión, una población sana contribuye con los impuestos. Esto representa flujos positivos de efectivo, y caso

contrario una población enferma requiere pagos e inversión en salud por parte del gobierno, generando un flujo negativo. De la misma manera las pensiones se ven afectadas, pues malos resultados en salud resultan en retiros a edades tempranas. Por lo que el uso de la medicina de precisión tendrá impacto directo en los flujos de efectivo. Es decir, una población sana genera contribuciones económicas positivas en términos de trabajo y disminución de gastos (Garrison and Towse 2017).

La medicina de precisión requiere de sistemas de determinación de precios, financiamiento y pago para demostrar su valor económico. Cuando se va a determinar el precio de un medicamento o tratamiento, las compañías farmacéuticas y los compradores deberán negociar durante las diferentes etapas de la cadena de valor. En el caso del desarrollo del medicamento, los compradores pueden proporcionar información valiosa acerca de los comparadores que probablemente utilicen (Garrison and Towse 2017).

Posterior al registro, se debe determinar el precio de acceso a mercado, considerando los diferentes niveles de pacientes y los esquemas escalonados y una vez en el mercado o en proceso de comercialización, se tendrá evidencia mediante el monitoreo del acerca del desempeño del fármaco en la vida real dando lugar a un parámetro para tomar decisiones acerca de incrementar o disminuir el precio (“The Economics of Precision Medicine” 2018).

Aunque el precio de acceso a mercado es un precio negociado, tiene su base en aspectos como costo de desarrollo y manufactura, tamaño de la compañía farmacéutica, tamaño del mercado potencial, estructura del mercado y efectividad relativa contra las otras alternativas terapéuticas. Sin embargo, los procesos de negociación y las decisiones de reembolso son diferentes entre cada país dando

como resultado amplias diferencias de precio y accesibilidad para el paciente (“The Economics of Precision Medicine” 2018).

11 Estado del arte de la medicina de precisión a nivel global

La medicina de precisión como modelo de salud pública es muy reciente sin embargo ya hay varios países que han empezado a implementarla. A continuación, se mencionarán los 3 casos más avanzados

GB Reino Unido

En Reino Unido (RU) la iniciativa de la medicina de precisión está pautada por *100,000 Genomes Project* (mencionado anteriormente en este trabajo). Este proyecto fue lanzado en 2012 a la vez que surge *Genomics England*, una organización fundada para organizar y coordinar el trabajo planeado de la secuenciación de 100,000 genomas de 70,000 pacientes del Sistema Nacional de Salud. *Genomics England* se configura como una empresa, pero es propiedad del Departamento de Salud, y está dirigida por un consejo que incluye a varios científicos destacados del Reino Unido. Los pacientes son reclutados a través de 13 centros designados por el *NHS Genomic Medical Center* los cuales están ubicados en toda Inglaterra junto con otras regiones de este país, como Escocia, Gales e Irlanda del Norte. A la fecha, este programa ha reportado más de 30,000 genomas secuenciados teniendo como objetivo final avanzar el Sistema de salud hacia la medicina de precisión.

El gobierno del RU se ha comprometido a financiar este programa hasta 2021 y la prolongación de la inversión o una nueva inversión estará determinada a través de la evaluación del beneficio de los resultados. Investigadores y académicos han sido los que se han visto mayormente beneficiados, pues pueden acceder a los datos de *100,000 Genomes Project* al unirse a uno o más consorcios conocidos como

Genomics England Clinical Interpretation Partnerships (GeCIPs, por sus siglas en inglés). Este consorcio se enfoca en dos áreas principales: cáncer y enfermedades raras, no obstante, a través del GeCIPs se pueden enfocar esfuerzos en algún aspecto específico de estas enfermedades.

Genomics England también ha facilitado las asociaciones con empresas privadas como *Illumina*, la cual se desempeña como el socio para la secuenciación, mientras que otras empresas, incluidas *Congenica* y *WuXi NextCode*, están involucradas en el desarrollo de herramientas para analizar e interpretar los datos de las secuenciaciones y los registros de salud asociados. La estrategia de *Genomics England* implica explorar la utilidad de diferentes plataformas bioinformáticas a través de una "fase de prueba", donde diferentes empresas trabajan con un subconjunto de datos de secuencia del proyecto para demostrar las capacidades y limitaciones de sus plataformas.

RU también realizó un proyecto denominado *Genetics Expert Network for Enterprises* (GENE, por sus siglas en inglés) que reunió a académicos, centros de medicina genómica del NHS y socios del sector privado de los sectores biotecnológico y farmacéutico. Su objetivo principal fue el desarrollo de nuevos diagnóstico y tratamientos basados en datos del *100,000 Genomes Project* para demostrar que los resultados también son compatibles con las necesidades de la industria.

LR Estados Unidos

Para este país y derivado de la *Iniciativa para el desarrollo de la medicina de precisión* anunciada en 2015, se desarrolló el programa *All of Us*, el cual es un repositorio y estudio de cohortes que planea inscribir a más de un millón de estadounidenses, también mencionado anteriormente en este documento.

Este programa ha desarrollado dos métodos para reclutar e inscribir participantes. En el primero, 8 organizaciones de proveedores de atención médica reclutan participantes de sus poblaciones de pacientes. En el segundo, las personas se ofrecerán como voluntarios para participar en línea y luego visitarán una ubicación contratada para recolectar muestras biológicas. La metodología para hacer que estos datos y las muestras biológicas estén disponibles para los investigadores aún se encuentra en desarrollo, pero esta iniciativa tiene el objetivo de que esta información esté disponible para los investigadores académicos y comerciales. El financiamiento del proyecto de medicina de precisión incluido este programa, lo determina el Congreso de los Estados Unidos como parte de su presupuesto general.

En este mismo país, existen otros esfuerzos para llevar a cabo el modelo de medicina de precisión como el denominado *Cancer Moonshot* que persigue objetivos tales como, mejorar la inmunoterapia, el fomento de la investigación sobre el cáncer infantil, la detección de tumores, la prevención y detección temprana del cáncer, abordar la resistencia a los medicamentos, así como la recolección e intercambio de datos de la enfermedad.

Otro ejemplo de MP en Estados Unidos, es el proyecto eMERGE, el cual comenzó como una red de bio repositorios, que recientemente se ha convertido en un programa que también se enfoca en brindar resultados de investigación genómica predictiva en la atención clínica. Otro prototipo basado en el modelo de MP es *el Clinical Sequencing Evidence-Generating Research Consortium* (CSER2, por sus siglas en inglés), el cual tiene como objetivo integrar las pruebas de secuenciación genómica en la práctica clínica rutinaria, incluso en el diagnóstico y tratamiento de enfermedades raras.

A pesar de todos estos ejemplos, ninguno de estos esfuerzos implica implementaciones completas del modelo de MP, pues el Programa de Investigación *All of Us*, la Red eMERGE y el Consorcio CSER2, solo involucran estrategias para devolver los hallazgos de la secuenciación genómica a los participantes de la investigación y sus proveedores de atención médica. Los objetivos de investigación son diseñados para observar cómo estos resultados afectan la atención clínica, sin considerar algunos otros elementos del modelo como la integración de datos obtenidos a través de otras tecnologías ómicas.

JP **Japón**

En 2014, se aprobaron dos leyes relativas a la política de innovación sanitaria: la *Ley de Promoción de la Política Sanitaria* y la *Ley de la Agencia Administrativa Independiente de la Agencia Japonesa para la Investigación y el Desarrollo Médicos (AMED, por sus siglas en inglés)*.

Headquarters for Healthcare Policy, proporciona un centro organizativo central para promover estratégicamente la innovación sanitaria, con el objetivo principal de "acelerar la traducción fluida y cooperativa de la investigación básica a la aplicación clínica" mediante la concesión de subvenciones para la investigación.

Los objetivos de estas organizaciones indican que la iniciativa de atención médica impulsada a nivel nacional en Japón incorpora la realización de la medicina genómica como uno de sus objetivos clave por ejemplo la creación en 2015 del Council for Realization of Genomic Medicine (CRGM) y aunque todavía no se ha abordado un sistema de salud de aprendizaje operativo, varias iniciativas para recopilar datos genómicos y conectarlos con las EHR, especialmente en relación con el cáncer.

Por su parte AMED está impulsando tres proyectos de investigación:

- a) “Programa-Plataforma de Promoción de la Medicina del Genoma” (un proyecto de 5 a 10 años para enfermedades comunes)
- b) “Programa para una base de datos integrada de información clínica y genómica” (un proyecto de 3 a 5 años para enfermedades raras, cáncer y otros con el Ministerio de Salud, Trabajo y Bienestar (MHLW))
- c) “Programa para la Promoción de la Plataforma de Descubrimiento de Fármacos basado en Genómica” (un proyecto de 3 años para implementación clínica con MHLW). AMED también estableció una política de intercambio de datos genómicos en 2016.

Como resultado, el intercambio de datos genómicos es cada vez más necesario para las bases de datos japonesas. Por ejemplo, la base de datos humana de NBCC está diseñada tanto con un sistema administrado/controlado como con un sistema de acceso abierto.

Varios biobancos a gran escala han recibido apoyo continuo, incluida la Organización de Megabancos Médicos de Tohoku (ToMMo, por sus siglas en inglés). Este programa ha recogido muestras de más de 150.000 personas sanas y ha analizado unos pocos miles de secuencias de alta cobertura del genoma completo. Este recurso se ha utilizado para estimar las frecuencias de las variantes patogénicas en la población japonesa y se está fomentando la utilización activa de este y otros biobancos.

Otra actividad relevante, la Iniciativa sobre enfermedades raras y no diagnosticadas (IRUD, por sus siglas en inglés), se lanzó en 2015 para “maximizar el beneficio de los análisis del genoma completo y del exoma completo para los pacientes que buscan un diagnóstico.

Varios elementos de estos esfuerzos están comenzando a llegar a la atención clínica. Por ejemplo, en 2016, los servicios nacionales de salud realizaron un esfuerzo ampliando el número de enfermedades raras para las que se cubren las pruebas genéticas de 36 condiciones a 72 condiciones.

Perspectivas en México

Uno de los retos más importantes para nuestro país es incorporar a la medicina de precisión como parte del modelo de cuidado de la salud lo cual pone de manifiesto las distintas áreas en las que aún existen oportunidades de mejora para poner en marcha esta práctica, como el uso de tecnologías ómicas, legislación en materia de bioética y protección de datos, conducción de los estudios clínicos, entre otros.

Uno ejemplo del interés en la implementación de este modelo en nuestro país, se ha dado en el *Instituto Nacional de Medicinas Genómica (Imegen)*, en donde se cuenta con un laboratorio que permite realizar secuenciación genómica, un centro de estudios éticos, legales y sociales, así como la apertura de foros de debate entre profesionales de la salud que permiten compartir sus investigaciones y hallazgos. Otro ejemplo en nuestro país, es el enfocado al uso de biomarcadores, donde se realizó un esfuerzo por seis compañías asociadas a la *Asociación Mexicana de Industrias de Investigación Farmacéutica (AMIFF)* para unificar sus programas de biomarcadores en la detección de cáncer de pulmón de células no pequeñas (NSCLC, *non small cell lung cancer*, por sus siglas en inglés) con el objetivo de mejorar el diagnóstico de este tipo de cáncer, permitiendo que con una sola biopsia, se puedan realizar la prueba para cinco diferentes biomarcadores moleculares.

Aunque gradualmente podemos ver la presencia de ciertos elementos de la medicina de precisión en la práctica clínica, aún no existe un centro de investigación en México

que reúna todos los elementos necesarios para realizar esta práctica de manera formal. Uno de los primeros desafíos para nuestro país, es la creación de una entidad, institución u órgano regulador, que identifique los elementos disponibles y establezca un modelo de trabajo que permita reunirlos. Esto implica una fuerte inversión en investigación e innovación como un requisito para una implementación exitosa. Aunque este proceso de implementación es complejo, pues mundialmente tampoco se ha definido por completo, dicha práctica será fundamental para tener una asistencia sanitaria eficiente, métodos modernos de cuidado de la salud y el control de los pacientes acerca de sus propios datos de salud, influyendo a su vez en el desarrollo económico en materia de salud.

12 Conclusiones

Este modelo de medicina basado en una clasificación de las enfermedades que toma en cuenta las características moleculares y las causas subyacentes de éstas, ha permitido estatificar a los pacientes en subpoblaciones más específicas y demostrado tener buenos resultados al ofrecer mejores diagnósticos, así como tratamientos dirigidos para los pacientes que se traducen en beneficios y menos reacciones adversas.

Muestra del interés en el desarrollo de la medicina de precisión, ha sido el constante surgimiento de proyectos a lo largo del mundo enfocados particularmente, en el impulso de la tecnología para llevarla a cabo, como lo son: tecnologías ómicas, secuenciación del genoma humano, el expediente clínico electrónico, biobancos, bases de datos, análisis de Big data y bioinformática, pudiéndose mencionar proyectos como *100,000 Genomes Project* o *All of us* que ya permiten el intercambio de la información colectada y con la cual se persigue promover la práctica de la

medicina de precisión. Así como instituciones que han incorporado el análisis de datos genómicos para identificar la causalidad de enfermedades o buscar tratamientos que hasta ahora no han sido determinados, por ejemplo, Instituto de Medicina Genómica Geisinger, Scripps Health, la Clínica Cleveland, el Centro Médico de la Universidad de Vanderbilt y el Colegio Médico de Wisconsin.

Por otra parte, en este trabajo se ha mencionado que la medicina de precisión mediante sus 4 principales características: Preventiva, predictiva, personalizada y participativa presenta considerables ventajas frente a la actual medicina reactiva donde podemos anticiparnos a la evolución de la enfermedad y anticipar la respuesta a un tratamiento siendo necesaria la participación de investigadores, pacientes, personal clínico y laboratorios, así como las entidades regulatorias.

También es de relevancia mencionar que este nuevo modelo ha sido posible debido al vertiginoso desarrollo de las tecnologías ómicas y otorgan la posibilidad de detectar los factores asociados con el desarrollo de enfermedades, y con ello poder comprender que intervenciones se pueden realizar.

El tiempo ha perfeccionado estas tecnologías, su evolución y desarrollo a permitido que los resultados sean cada vez más precisos, así como asequibles económicamente, ejemplos de proyectos exitosos que incluyen el uso de estas herramientas son: ENCODE (*Encyclopedia of DNA Elements*, por sus siglas en inglés) ó *Global Alliance for Genomics and Health (GA4GH)*.

De manera general, pudieramos concluir que se presentarán multiples beneficios para cada uno de los actores implicados en la implementación de la MP, es decir, se trabaja en la construcción de un modelo que beneficia a todas las partes, por ejemplo: los ciudadanos: Se tornarán empoderados y mejor informados al tener acceso y control

de su información, así como facilitarles el acceso a fuentes médicas confiables y comprensibles. Los proveedores de la salud tendrán acceso a la información generada durante la investigación, así como en el ámbito clínico diario con el objetivo de utilizarla en el modelo de medicina de precisión asumiendo la responsabilidad que esto conlleva de hacer uso responsable de estos datos. Los profesionales de la salud irán adquiriendo una formación interdisciplinaria acorde con esta nueva práctica. Así como los médicos e investigadores trabajarán en conjunto para el desarrollo e implementación de la medicina de precisión, en el caso de los Sistemas de **salud** permitirán la promoción de la salud y prevención, diagnóstico y tratamiento optimizados y personalizados para el beneficio de los pacientes. Los servicios de PM están optimizados en términos de efectividad y equidad. Asegurar el acceso de los datos de salud de los ciudadanos a los sistemas de salud y las autoridades reguladoras para alimentar las bases de datos e incrementar la investigación relacionada con la salud, Y para la Información clínica: Los proveedores de atención médica y los investigadores utilizarán de manera cotidiana los datos de los registros médicos electrónicos (EHR) para lograr una MP, alineada con las medidas necesarias que permitan garantizar la privacidad, la seguridad y la protección de los datos en la gestión de datos de salud.

Aun queda mucho por recorrer respecto a la medicina de precisión, como disciplina relativamente joven, es necesario que cada vez se difunda sobre ella y sobre los diferentes elementos que la componen, ya que en el futuro es más frecuente que los médicos y el resto de la población se encuentre más familiarizada con los términos y las nuevas pruebas que se realizan para diversas enfermedades y así mismo al volverse frecuente su uso, el costo será más accesible y las pruebas necesarias serán

mas frecuentemente utilizadas e incluidas en bases de datos electrónicas de manera regular.

Sobre este mismo tenor, no se debe dejar de trabajar en la legislación que protege los datos y la integridad de los pacientes o individuos que comparten información o participan en estudios clínicos, de ahí la necesidad constante de la sensibilización sobre las regulaciones vigentes a todos los involucrados en el modelo de medicina de precisión.

En lo referente a la capacitación de personal, es necesario la formación de profesionales que dominen conceptos de biología molecular, así como de programación computacional para la creación de algoritmos y programas estadísticos armonizados que puedan procesar y analizar, todo esto con el fin de que los profesionales de la salud puedan consultar resultados procesados de las investigaciones y sean capaces de identificar tendencias cuyo desenlace sea en beneficio del paciente, por lo que sería interesante la creación de una formación específica de profesionales que se dediquen exclusivamente al desarrollo de la medicina de precisión.

En conclusión, la medicina de precisión es un modelo exitoso e innovador de mejora en el cuidado clínico, sin embargo, aún queda mucho camino por recorrer para los pioneros en esta nueva práctica, cada una de las partes involucradas evoluciona a diferentes velocidades y su integración en sólo modelo es el reto al que todos nos enfrentamos.

13 Bibliografía

- Aronson, Samuel J., and Heidi L. Rehm. 2015. "Building the Foundation for Genomics in Precision Medicine." *Nature* 526 (7573): 336–42.
- Ashley, Euan A. 2016. "Towards Precision Medicine." *Nature Reviews. Genetics* 17 (9): 507–22.
- Bauch, Angela, Izabela Adamczyk, Piotr Buczek, Franz-Josef Elmer, Kaloyan Enimanev, Pawel Glyzowski, Manuel Kohler, et al. 2011. "OpenBIS: A Flexible Framework for Managing and Analyzing Complex Data in Biology Research." *BMC Bioinformatics* 12 (1): 468.
- Cecchin, Erika, and Gabriele Stocco. 2020. "Pharmacogenomics and Personalized Medicine." *Genes* 11 (6): 679.
- Chen, Jiajia, Yuxin Lin, and Bairong Shen. 2017. "Informatics for Precision Medicine and Healthcare." *Advances in Experimental Medicine and Biology* 1005: 1–20.
- "Clinical Decision Support." n.d. Accessed February 14, 2022. <https://www.healthit.gov/topic/safety/clinical-decision-support>.
- Doerr, Megan and Suver, Christine and Wilbanks, John., ed. 2016. *Developing a Transparent, Participant-Navigated Electronic Informed Consent for Mobile-Mediated Research*. Sage Bionetworks.
- Doerr, Megan, Christine Suver, and John Wilbanks. 2016. "Developing a Transparent, Participant-Navigated Electronic Informed Consent for Mobile-Mediated Research." *SSRN Electronic Journal*. <https://doi.org/10.2139/ssrn.2769129>.
- "Electronic Medical Records and Genomics (EMERGE) Network." n.d. Genome.Gov. Accessed March 14, 2022. <https://www.genome.gov/Funded-Programs-Projects/Electronic-Medical-Records-and-Genomics-Network-eMERGE>.
- Garrison, Louis P., and Adrian Towse. 2017. "Value-Based Pricing and Reimbursement in Personalised Healthcare: Introduction to the Basic Health Economics." *Journal of Personalized Medicine* 7 (3): 10.
- Glicksberg, Benjamin S., Kipp W. Johnson, and Joel T. Dudley. 2018. "The next Generation of Precision Medicine: Observational Studies, Electronic Health Records, Biobanks and Continuous Monitoring." *Human Molecular Genetics* 27 (R1): R56–62.
- Hood, Leroy, and Lee Rowen. 2013. "The Human Genome Project: Big Science Transforms Biology and Medicine." *Genome Medicine* 5 (9): 79.
- Kinkorová, Judita. 2015. "Biobanks in the Era of Personalized Medicine: Objectives, Challenges, and Innovation: Overview." *The EPMA Journal* 7 (1): 4.
- Manzoni, Claudia, Demis A. Kia, Jana Vandrovцова, John Hardy, Nicholas W. Wood, Patrick A. Lewis, and Raffaele Ferrari. 2018. "Genome, Transcriptome and Proteome: The Rise of Omics Data and Their Integration in Biomedical Sciences." *Briefings in Bioinformatics* 19 (2): 286–302.
- Martin-Sanchez, Fernando, Riccardo Bellazzi, Vittorio Casella, William Dixon, Guillermo Lopez-Campos, and Niels Peek. 2020. "Progress in Characterizing the Human Exposome: A Key Step for Precision Medicine." *Yearbook of Medical Informatics* 29 (1): 115–20.
- McGrath, Scott, and Dario Gherzi. 2016. "Building towards Precision Medicine: Empowering Medical Professionals for the next Revolution." *BMC Medical Genomics* 9 (1). <https://doi.org/10.1186/s12920-016-0183-8>.
- Olivier, Michael, Reto Asmis, Gregory A. Hawkins, Timothy D. Howard, and Laura A. Cox. 2019. "The Need for Multi-Omics Biomarker Signatures in Precision

- Medicine." *International Journal of Molecular Sciences* 20 (19): 4781.
- Primorac, Dragan, Lidija Bach-Rojecky, Dalia Vađunec, Alen Juginović, Katarina Žunić, Vid Matišić, Andrea Skelin, et al. 2020. "Pharmacogenomics at the Center of Precision Medicine: Challenges and Perspective in an Era of Big Data." *Pharmacogenomics* 21 (2): 141–56.
- Pushpakom, Sudeep, Francesco Iorio, Patrick A. Eyers, K. Jane Escott, Shirley Hopper, Andrew Wells, Andrew Doig, et al. 2019. "Drug Repurposing: Progress, Challenges and Recommendations." *Nature Reviews. Drug Discovery* 18 (1): 41–58.
- Quezada, Héctor, Ana Laura Guzmán-Ortiz, Hugo Díaz-Sánchez, Ricardo Valle-Rios, and Jesús Aguirre-Hernández. 2017. "Omics-Based Biomarkers: Current Status and Potential Use in the Clinic." *Boletín Médico Del Hospital Infantil de México (English Edition)* 74 (3): 219–26.
- Samperi, Tena Inés. 2020. "Estrategias de Identificación de Variantes Asociadas a Caracteres Productivos Mediante Chips o Secuenciación a Baja Cobertura." Master Ingeniería Agronómica, España: Universidad de Lleida.
- Sandhu, Charanjit, Alia Qureshi, and Andrew Emili. 2018. "Panomics for Precision Medicine." *Trends in Molecular Medicine* 24 (1): 85–101.
- "The Economics of Precision Medicine." 2018. presented at the IRGC Expert Workshop, Campus Biotech, Geneva, April 12.
<https://www.epfl.ch/research/domains/irgc/specific-risk-domains/projects-precision-medicine/page-154106-en-html/>.
- Uozie, Anuli Christiana, and Ruedi Aebersold. 2018. "Advancing Translational Research and Precision Medicine with Targeted Proteomics." *Journal of Proteomics* 189 (October): 1–10.
- Wang, Zhining, Mark A. Jensen, and Jean Claude Zenklusen. 2016. "A Practical Guide to the Cancer Genome Atlas (TCGA)." *Methods in Molecular Biology (Clifton, N.J.)* 1418: 111–41.
- Wang, Zhong, Mark Gerstein, and Michael Snyder. 2009. "RNA-Seq: A Revolutionary Tool for Transcriptomics." *Nature Reviews. Genetics* 10 (1): 57–63.
- Weichenhan, Dieter, Daniel B. Lipka, Pavlo Lutsik, Ashish Goyal, and Christoph Plass. 2020. "Epigenomic Technologies for Precision Oncology." *Seminars in Cancer Biology*, August. <https://doi.org/10.1016/j.semcancer.2020.08.004>.
- Wishart, David S. 2016. "Emerging Applications of Metabolomics in Drug Discovery and Precision Medicine." *Nature Reviews. Drug Discovery* 15 (7): 473–84.