

ANEXO I. ESPECIFICACIONES TÉCNICAS PARA LA BÚSQUDA DE SOLUCIONES DE INNOVACIÓN EN EL SISTEMA SANITARIO PÚBLICO DE ANDALUCÍA UNA CONSULTA PRELIMINAR DE MERCADO PARA LA BÚSQUDA DE SOLUCIONES DE INNOVACIÓN EN EL SISTEMA SANITARIO PÚBLICO DE ANDALUCÍA PARA EL DESARROLLO DE UN SISTEMA INTEGRAL DE MANEJO DE DATOS GENÓMICOS PARA LA MEDICINA PERSONALIZADA EN ANDALUCÍA “MedPre_Andalucía”

El contenido del presente documento describe con carácter orientativo la necesidad no cubierta, la cual podrá evolucionar a medida que se vaya actualizando el estado del arte para este proyecto.

1. Antecedentes

Andalucía cuenta con uno de los mayores sistemas sanitarios de Europa, con una extensión mayor que 30 países europeos y ofreciendo asistencia a una población de 8,4 millones de habitantes, un 18% del total nacional.

El Sistema Sanitario Público de Andalucía (SSPA), además de dar respuesta a los problemas de salud de los andaluces, desempeña un importante rol en I+i, resultando una pieza clave y dinamizadora de la competitividad de la comunidad, ya que es el pilar básico de la detección de las necesidades de los pacientes, se trata de una de las mayores estructuras generadoras de conocimiento, dispone de las infraestructuras y el *expertise* necesario para ofrecer servicios acordes con las demandas industriales y sociales, y se trata de una organización con capacidad de generar un retorno económico y social.

A nivel regional, la Estrategia en I+i en Salud 2020-2023 se vertebra en 5 líneas maestras, entre las que destacan:

- La consolidación del valor de la I+i en salud por y para la sociedad, mediante el desarrollo de un conocimiento de excelencia.
- Posicionar la I+i de Andalucía a nivel nacional e internacional.
- Estimular y facilitar la colaboración público-privada.
- Potenciar la I+i en genómica, terapias avanzadas y enfermedades raras, junto al desarrollo de la medicina móvil, macrodatos (big data) e inteligencia artificial, con el objetivo de avanzar en la medicina personalizada.

Los datos genómicos constituyen el paradigma de datos voluminosos y complejos, de manejo complicado, pero con un gran potencial para su uso clínico. Es por ello que la generación de software de manejo de los datos es un campo en constante evolución. El fin último de esta propuesta es convertir a Andalucía en un referente nacional e internacional en materia de desarrollo de Big Data clínico, unido a aplicaciones de Inteligencia Artificial (IA), los cuales son considerados pilares de la Medicina de Precisión o P4 (Personalizada, Predictiva, Preventiva y Participativa).

La necesidad de este proyecto de Compra Pública de Innovación (CPI) viene generada por la demanda de una medicina de precisión, favorecida por el abaratamiento de las metodologías de secuenciación, que permiten la detección de biomarcadores diagnósticos, pronósticos, de recomendación de tratamiento y preventivos. El manejo actual está basado en soluciones



comerciales muy parciales, basadas en paneles de genes y con un software no integrado, ni integrable en los sistemas de salud y que no cubren más que algunas necesidades diagnósticas y algunas posibilidades de tratamiento en cáncer. Hay empresas que ofrecen soluciones hospitalarias con manejo integral, pero solo de datos genómicos de enfermedades genéticas y cáncer, sin posibilidad de un desarrollo real. No existe ninguna solución centralizada y con un ámbito tan amplio e integral, como la que aquí se propone.

Andalucía ha sido una comunidad pionera al ser la primera que tuvo la visión de hacer un proyecto de secuenciación genómico orientado a la clínica: el Proyecto del Genoma Médico (MGP) [1], un proyecto que fue llevado a cabo en 2011, mucho antes que otras iniciativas en otras comunidades. Posteriormente, otras comunidades han lanzado sus propios proyectos, como el proyecto Nagen [2], para secuenciar los genomas completos de 1000 pacientes de enfermedades raras y cánceres, iniciado en 2017 (cuyo análisis de datos se subcontrató, precisamente, a la antigua Área de Bioinformática, ahora parte de la Plataforma de Medicina Computacional (PMC) de Andalucía). Cataluña ha financiado mediante sus proyectos PERIS del 2017 la iniciativa URDCat [3] para explorar la capacidad de diagnóstico que la secuenciación tiene para las enfermedades raras. Madrid ha lanzado en 2018 su iniciativa RareGenomics [4] con un objetivo similar. Y a nivel nacional, el Centro de Investigación Biomédica en Red de Enfermedades Raras (CIBERER) lanzó en 2015 la segunda fase del proyecto de Enfermedades no Diagnosticadas (EnoD) [5] para descubrir las bases genéticas de enfermedades raras en pacientes a los que no se pudo dar un diagnóstico con el conocimiento actual. Como grupo miembro del CIBERER [6], la PMC participa en el análisis de los datos genómicos. En ninguna de estas iniciativas existe ningún plan para la gestión posterior de los datos genómicos. En varias de ellas (Navarra, Cataluña, CIBERER) los datos salen del sistema sanitario con muy escasa información clínica (principio de la mínima información según LOPD) con lo que su aprovechamiento para generar conocimiento más allá del proyecto es prácticamente imposible.

Más recientemente, el ISCIII ha lanzado el proyecto Infraestructura de Medicina de Precisión asociada a la Ciencia y la Tecnología -IMPACT [7], en cuyo pilar IMPACT-datos, se coordina desde la PMC el paquete de trabajo de integración de datos. Desde IMPACT-data estamos exportando la experiencia de Andalucía en el manejo de datos clínicos al sistema de salud nacional.

En la actualidad, la secuenciación en los distintos hospitales se hace de forma descoordinada con secuenciadores de relativa poca procesividad, con lo que la secuenciación resulta más cara. Aun así, se han puesto en marcha dos proyectos que pueden ser icónicos para la implantación de la medicina personalizada. El primero es un proyecto piloto para el diagnóstico de pacientes con enfermedades genéticas puesto a punto en el Hospital Universitario Virgen del Rocío [8], mediante el cual se han hecho más de 5.000 diagnósticos por el personal del servicio de Genética Humana y Reproducción, sin necesidad de intervención de personal especializado bioinformático, lo que garantiza su escalabilidad a otros servicios y hospitales, así como la centralización de los datos generados. De hecho, durante la pandemia de la COVID-19 se puso en marcha un Circuito de Vigilancia Genómica [9], que implica la recolección coordinada de muestras de Andalucía Oriental en el Hospital San Cecilio (HUSC) y de Andalucía Occidental en el Hospital Virgen del Rocío (HUVR),



empresas, las cuales proporcionan programadores que participan en el desarrollo del software y en este proceso adquieren conocimiento, que les permite luego ofrecer instalaciones personalizadas y un soporte profesional a los sistemas de salud.

Andalucía también tiene mucho camino recorrido con respecto al uso secundario de datos. La Base Poblacional de Salud (BPS) [\[12\]](#), con sus 13 millones de pacientes, es el recurso más grande de España de estas características y probablemente del mundo. Con esta idea de cambio de paradigma en el sistema de salud, desde la PMC se ha promovido una experiencia piloto de entorno de investigación de confianza, siguiendo la filosofía de Genomics England (el proyecto de los 100.000 genomas de UK, donde pusieron en marcha el Trusted Research Environment [\[13\]](#), que consiste en un entorno computacional donde los datos protegidos por RGPD pueden ser estudiados dentro del sistema de salud, minimizando así los riesgos de pérdida de datos o re-identificación de pacientes. La Plataforma para la Generación segura de Conocimiento a partir del Big Data Clínico de la BPS (iRWD) [\[14\]](#) es un entorno de investigación de confianza consistente en una infraestructura computacional localizada dentro de la red corporativa y accesible desde BPS que fue adquirida mediante un proyecto de “Ayudas a infraestructuras y equipamientos de I+D+i, en la modalidad adquisición de material científico y mejora de infraestructuras I+D+i, para entidades de carácter público convocada en 2019, en régimen de concurrencia competitiva, en el ámbito del Plan Andaluz de Investigación, Desarrollo e Innovación (PAIDI 2020, IE19_259 FPS)”, En dicha infraestructura se han realizado ya varios proyectos financiados tanto por fondos competitivos como por empresas.

Finalmente, el Programa Andaluz de Formación en Medicina Personalizada y de Precisión (PANMEP) es un ejemplo de iniciativa de formación a nivel regional, que ha formado ya a una primera generación de profesionales del SAS en diversos aspectos y metodologías relacionadas con la implantación de la medicina de precisión en el sistema de salud y seguirá formando en el futuro.

2. Necesidad no cubierta

La digitalización de los sistemas de salud, junto con la adopción de nuevas tecnologías, están generando datos clínicos en cantidades masivas y a un ritmo nunca antes visto. Los sistemas de salud se están convirtiendo en uno de los mayores generadores de datos de nuestro tiempo, pudiéndose hablar, con toda propiedad, del Big Data clínico.

El Big Data clínico, junto con las aplicaciones de la IA en biomedicina y salud, pueden considerarse los principales pilares de la Medicina de Precisión o Medicina P4: Personalizada, Predictiva, Preventiva y Participativa. Sin embargo, cuando se habla de ella, siempre se alude a la personalización de los tratamientos o del diagnóstico, que correspondería a un uso primario de la información genómica, olvidando, pues, otras facetas que ofrece la generación de conocimiento mediante el uso secundario de estos datos, como son:

- La predictiva, para predecir el curso de la enfermedad o necesidades de medicación.
- La preventiva, adelantándose a la enfermedad y permitiendo tomar medidas.



- La participativa, en la que el propio paciente toma protagonismo en la gestión integral de su salud.

En este escenario, el uso de los datos genómicos en la práctica clínica rutinaria requiere una estrategia completamente distinta a la actual, en la que se incluyan un alto grado de automatización y un software que permita una trazabilidad integral de los pacientes, sus datos genómicos, sus resultados y las decisiones derivadas, incluyendo sistemas de alertas.

La finalidad de este proyecto se dirige, por tanto, a romper con el esquema tradicional de manejo de datos genómicos, basado en:

- a. La utilización de equipos bioinformáticos que realizan gran parte del análisis de los datos de forma manual, lo cual implica tiempos de respuesta mínimos entre 15-16 semanas. Asimismo, por regla general, la generación de informes y su inclusión en la Historia de Salud Digital se realiza de forma manual, en formatos no estructurados como, por ejemplo, en formato PDF, sobre los que no es posible efectuar consultas electrónicas.
- b. El uso primario de la información genómica, olvidando el uso secundario -pero no menos importante- de estos datos, antes descrito.

Actualmente, no hay empresas con conocimiento previo de estas aplicaciones que puedan ofrecer los servicios de instalación en producción, integración con los sistemas corporativos, mantenimiento y soporte ininterrumpido.

Dando respuesta a la necesidad planteada, se conseguirá permitir la secuenciación genómica de pacientes -biopsias, patógenos infecciosos o ambientales y microbiota intestinal- y utilizar la información obtenida para el diagnóstico, la recomendación de tratamiento y el seguimiento de los pacientes.

Además, permitirá a los pacientes interoperar con su información genómica de forma pionera para tomar decisiones informadas preventivas o sobre estilo de vida. Es por ello que uno de los aspectos con mayor innovación de esta plataforma es la posibilidad de transformar datos genómicos no estructurados y de difícil uso, en datos estructurados para facilitar el uso secundario de esta información, junto con el resto del Big data clínico existente en la Base Poblacional de Salud (BPS), de forma que acelere el descubrimiento de biomarcadores y aumente la precisión de la práctica médica en el sistema de salud. Todo ello, redundará en una mayor calidad en la atención del paciente y una reducción de los costes de tratamientos y hospitalizaciones, permitiendo, además, una vez validada, su implementación a otros sistemas de salud.

3. Estado del Arte

Digitalización y Big data clínico

La paulatina digitalización de los sistemas de salud, junto con la adopción de nuevas tecnologías, entre las que se encuentran la genómica, la imagen médica y más recientemente los biosensores y la telemedicina, están generando datos de carácter clínico en cantidades



masivas a un ritmo que nunca antes había sido conocido. Una de las consecuencias de esta transformación es que los sistemas de salud se están convirtiendo en uno de los mayores generadores de datos de nuestra sociedad, pudiéndose ya hablar con toda propiedad del Big Data clínico. En este escenario, la capacidad de analizar datos biomédicos a la antigua usanza ha sido claramente sobrepasada por el volumen actual y la forma de interpretarlos necesita cada vez más de computación y software especializado [15]. Esto hace que dentro de los sistemas de salud el número de decisiones dependientes de sistemas informáticos y de sistemas de apoyo a la decisión esté creciendo enormemente. Sin embargo, detrás de este aumento de la complejidad de los datos y de su procesamiento se esconde una oportunidad. Estos datos, producto de la digitalización de los sistemas de salud, pueden ser usados para mejorar los procesos del sistema sanitario, para entender de forma retrospectiva cuáles son las claves del éxito o el fracaso de tratamientos, etc. De hecho, con el crecimiento de las bases de datos en todos los sectores han florecido numerosas aplicaciones de metodologías de inteligencia artificial (IA), especialmente en el campo de la salud [16]. La IA está empezando a integrarse en distintos aspectos de la medicina, con la perspectiva de optimizar procesos, diagnósticos y tratamientos, así como disminuir errores médicos [17].

En general, el Big Data clínico, junto con las aplicaciones de la IA en biomedicina [18] y salud [19], pueden considerarse los principales pilares de la medicina de precisión, o medicina P4 (por Personalizada, Predictiva, Preventiva y Participativa) [20]. Sin embargo, cuando se habla de medicina de precisión siempre se alude a la personalización de tratamientos o del diagnóstico, que correspondería al uso primario de la información genómica, olvidando otras muchas facetas que ofrece la generación de conocimiento mediante el uso secundario de estos datos, como son la predictiva, con posibilidad de predecir el curso de la enfermedad o necesidades de medicación, la preventiva, adelantándose a la enfermedad y permitiendo tomar medidas, y finalmente la participativa, en la que el propio paciente toma protagonismo en la gestión integral de su salud, usando información derivada de sus datos genómicos y clínicos para optimizar sus tratamientos (evitando efectos adversos o secundarios) y modificar sus hábitos de vida (alimenticios, de ejercicio, etc.) y proporcionando datos transversales de su salud (microbioma). La razón principal por la que 3 de las P de la medicina P4 no llegan al paciente está precisamente en las dificultades de procesamiento, manejo y almacenamiento del Big data clínico, especialmente el genómico, que esta iniciativa pretende solucionar.

El ciclo de los datos en medicina de precisión

Realmente, la medicina de precisión es el resultado de la creciente generación de conocimiento debido al incremento en la producción de datos que provee un sustrato cada vez mayor para la investigación clínica. Es por ello que se habla del ciclo de datos en medicina de precisión. El uso primario de los datos que se extraen de los pacientes para su manejo (su diagnóstico, la recomendación de tratamientos, el pronóstico de su enfermedad, etc.) está basado en el conocimiento previo que permite interpretar en clave clínica el dato médico al paciente (dato genómico, imagen médica, analítica, etc.). Pero, ¿de dónde viene ese conocimiento? Precisamente vienen del estudio de esos datos que se generan para el manejo del paciente



(Figura 1).

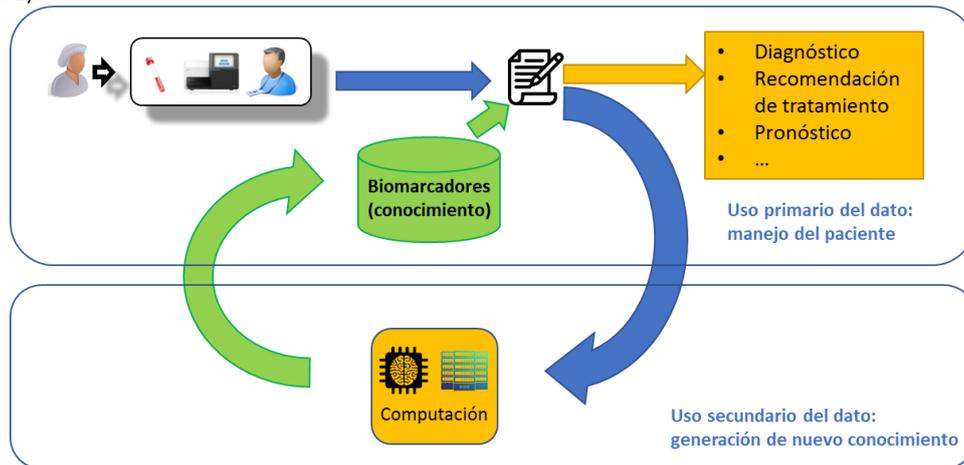


Figura 1. El ciclo de los datos en medicina de precisión

Aunque en el pasado, cuando el sistema de salud no estaba aún digitalizado, el proceso de generación del conocimiento biomédico era completamente externo al sistema de salud y los datos se producían ex profeso como parte de este proceso, la digitalización del sistema de salud ha cambiado radicalmente el escenario y en estos momentos este se ha convertido en el mayor generador de datos biomédicos.

Hablamos de datos del mundo real (RWD, del inglés *Real World Data*) en medicina para referirnos a datos derivados de una serie de fuentes que se asocian con los resultados en una población heterogénea de pacientes en entornos del mundo real, como encuestas a pacientes, ensayos clínicos y estudios de cohortes observacionales. Los datos del mundo real se refieren a los datos observacionales en contraposición a los datos recogidos en un entorno experimental como un ensayo clínico. Y a partir del RWD hablamos de evidencia del mundo real (RWE del inglés *Real World Evidence*) o evidencia obtenida de los datos del mundo real (RWD).

Sistema de salud productor de datos pasivo o generador de conocimiento

En la actualidad, el sistema de salud es uno de los mayores generadores de datos de nuestra sociedad, mostrando una producción de datos clínicos prácticamente exponencial. Se ha estimado que hasta 1950 se necesitaban más de 50 años para que el volumen de datos médicos se duplicase, pero esta tasa de generación de conocimiento se aceleró a solo 7 años en 1980, a 3 años en 2011, y se estima que en 2020 ocurrió cada 73 días [21]. Esto convierte al sistema de salud en una fuente enorme de RWD a partir del cual, por RWE se puede generar nuevo conocimiento clínico, que a su vez contribuirían a una medicina más precisa (ver en la Figura 1 el ciclo de los datos en medicina de precisión).

Paradójicamente, el escenario actual para usar la RWE está absolutamente obsoleto y pertenece a una época anterior a la digitalización de los hospitales. Por ejemplo, los centros de cálculo y



los grandes recursos computacionales (el Centro Nacional de Supercomputación de Barcelona, o el CICA andaluz, etc.) están normalmente desvinculados del sistema hospitalario, en redes informáticas distintas, y pertenecen a distintas administraciones. Sin embargo, es el sistema de salud el que hace una enorme inversión en la generación de los datos y además tiene la experiencia de sus profesionales sobre como extraer conocimiento biomédico de ellos.

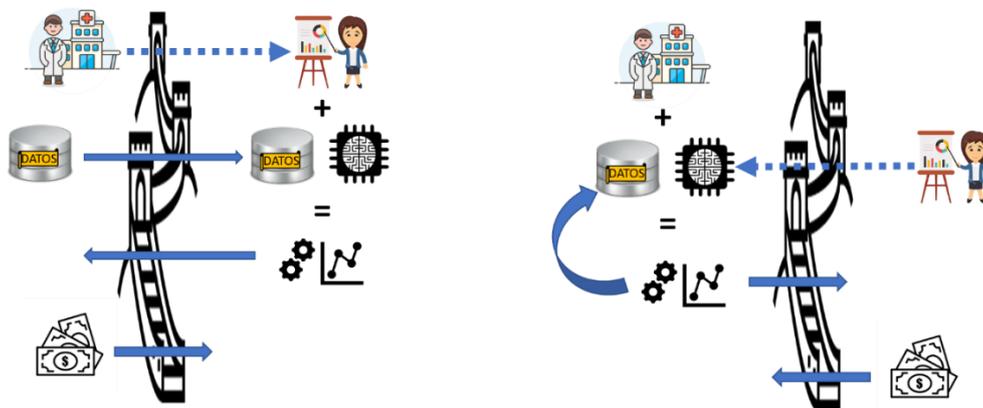


Figura 2. Izquierda: situación actual, en la que el sistema de salud es productor de datos y para su uso secundario se extraen anonimizados, se realiza el estudio por agentes externos (con la ayuda de clínicos muchas veces) y se genera conocimiento, por el que a veces hay que pagar para implementar en el sistema de salud. Derecha: cambio de paradigma en la era de la post-digitalización del sistema de salud. Se dota al sistema de salud de capacidad de computación y el conocimiento se genera dentro, con la colaboración de agentes externos, y en caso de que ese conocimiento sea comercializable, contribuye la sostenibilidad del sistema de salud.

En la Figura 2 esquematiza cómo se entiende el uso secundario de datos en la actualidad: extracción de datos, análisis en el exterior y los potenciales resultados de interés para el sistema de salud (ej. predictores, sistemas de soporte a la decisión, etc.) han de ser adquiridos por este, normalmente con un coste asociado. En resumidas cuentas: el sistema de salud genera los datos, aporta el conocimiento para su análisis y acaba pagando por los desarrollos de esa investigación. La parte derecha esquematiza el cambio de paradigma propuesto: el sistema de salud se dota de capacidad de computación y con su experiencia (y la experiencia analítica externa si hace falta) realiza su propia investigación clínica, cuyos desarrollos son implementables sin coste en el sistema de salud, y en caso de ser comercializables, el ingreso resultante contribuiría al sostenimiento del sistema de salud.

4. Objetivos del proyecto

El objetivo general de este proyecto consiste en el diseño y desarrollo una plataforma integral de uso de datos genómicos en Medicina de Precisión, la cual permitirá un manejo mucho más preciso del paciente, con mayor exactitud en los diagnósticos y decisiones de tratamientos mucho más eficaces, lo que redundará en una mayor calidad en la atención del paciente y una reducción de los costes de tratamientos y hospitalizaciones.



Como aspectos más innovadores del proyecto, se destacan las siguientes funcionalidades de la plataforma:

- El uso primario de datos genómicos para el manejo del paciente a nivel hospitalario en los servicios que lo usen, inicialmente genética, patología e infecciosos, pero, posteriormente, otros servicios.
- El uso secundario a partir de una base de datos centralizada, que favorecerá la generación de conocimiento biomédico, aumentando la posibilidad de prevención en la práctica médica, frente a la medicina actual de carácter más reactivo.

La solución tendrá un carácter regional (Andalucía), pudiendo, una vez validada, ser implementada a otros sistemas de salud.

Entre los objetivos específicos que presenta este proyecto se encuentran los siguientes:

- 1) Dotar al sistema de salud de la capacidad de secuenciación genómica a gran escala para su uso en la prestación de servicios sanitarios.
- 2) Dotar al sistema de salud de la capacidad de manejar los datos genómicos para su uso primario -diagnóstico, pronóstico y de decisión de tratamiento- de forma integral.
- 3) Ampliar el alcance del uso de la secuenciación genómica a los patógenos ambientales con incidencia en la salud.
- 4) Ampliar el uso de la secuenciación genómica a la microbiota de los pacientes.
- 5) Proporcionar a los pacientes la posibilidad de interoperar con información derivada de sus datos genómicos para tomar decisiones informadas preventivas o sobre estilo de vida.
- 6) Dotar al sistema de salud de un entorno de investigación de confianza interoperable desde el exterior.
- 7) Dotar al sistema de salud de la capacidad innovadora de usar datos genómicos como datos estructurados junto con otros datos clínicos para el uso secundario de los mismos en investigación.

5. Servicios interesados en desplegar

Atendiendo a los objetivos expuestos anteriormente, MedPre_Andalucia requiere del desarrollo de las siguientes funcionalidades:

- a. **Sistema centralizado de pre-procesamiento de datos genómicos:** Permitirá una descarga directa de los datos de los equipos de secuenciación a un repositorio único donde, de forma automática, permitiría su control de calidad y transformación en datos usables en la práctica clínica. Permitirá su portabilidad a arquitecturas altamente paralelas para su escalado a un uso regional. En relación con los niveles de madurez tecnológica, el TRL actual es el 7 y, el que se pretende obtener será un TRL8. Los riesgos tecnológicos, que comportaría esta



funcionalidad se basarían en la complejidad para la portabilidad de algunos algoritmos a arquitecturas altamente paralelas.

- b. Sistema de manejo de datos genómicos para su uso primario en diagnóstico de enfermedades hereditarias:** Permitirá un uso directo del dato genómico para el diagnóstico de enfermedades hereditarias. Paneles virtuales. No se considera necesario re-secuenciar para diagnósticos posteriores. Hallazgos secundarios (actualizable). Perfil farmacogenómico. No existen soluciones orientadas a la práctica clínica. Apenas existen algunas aplicaciones muy orientadas a la investigación en diagnóstico de enfermedades raras. Actualmente, no existen soluciones comerciales. Existe un proyecto piloto en el Hospital Universitario Virgen del Rocío (HUVR) en el que se han hecho más de 5.000 diagnósticos. El TRL actual es el 7 y, el que se pretende obtener será un TRL8. Respecto a los riesgos tecnológicos, podrían plantearse problemas de escalabilidad y de extensión a distintos hospitales dependiendo del ancho de banda disponible.
- c. Sistema de manejo de datos genómicos para su uso primario en decisión de tratamiento en cáncer:** Permitirá un uso directo del dato genómico para la decisión de tratamiento en cáncer en función del perfil mutacional de la biopsia. Perfil farmacogenómico. Actualmente, no existen soluciones orientadas a la práctica clínica. Apenas existen algunas aplicaciones y las existentes se encuentran muy orientadas a la investigación. Asimismo, no existen soluciones comerciales disponibles en el mercado. El TRL actual es el 4 y, el que se pretende obtener será un TRL8. Respecto a los riesgos tecnológicos, podrían plantearse problemas de escalabilidad y de extensión a distintos hospitales dependiendo del ancho de banda disponible.
- d. Sistema de manejo de datos genómicos de patógenos virales y bacterianos para el manejo de pacientes infecciosos:** Permitirá un uso directo del dato genómico de virus y bacterias para el control epidemiológico y la decisión de tratamiento en los casos de resistencias a antivirales (virus) o a antibióticos (bacterias). Actualmente, no existen soluciones orientadas a la práctica clínica. Apenas existen algunas aplicaciones muy orientadas a la investigación. Asimismo, no existen soluciones comerciales. Actualmente, se ha desarrollado una aplicación piloto en el HUVR denominada Circuito de vigilancia genómica. El TRL actual es el 5 y, el que se pretende obtener será un TRL8. Respecto a los riesgos tecnológicos, podrían plantearse problemas con la trazabilidad de las muestras si no se realiza una correcta integración en los sistemas de petición de análisis.
- e. Sistema de vigilancia epidemiológico basado en datos genómicos de patógenos virales ambientales:** Permitirá un uso directo de datos genómicos de virus y bacterias para el control epidemiológico. Actualmente, no existen soluciones orientadas a la práctica clínica. Actualmente, apenas existen algunas aplicaciones y, las existentes, se encuentran muy orientadas a la investigación. Actualmente, no existen soluciones comerciales para esta necesidad. La aplicación SIEGA ha sido desarrollada como piloto en el HUVR. El TRL actual es el 5 y, el que se pretende obtener será un TRL8. Respecto a los riesgos tecnológicos, podrían plantearse problemas con la trazabilidad de las muestras si los



códigos de las muestras externas no se integran correctamente con los códigos de los sistemas de petición de análisis.

- f. Sistema de manejo de datos genómicos de la microbiota para la toma de decisiones y la prevención con los pacientes:** Permitirá un uso totalmente innovador del dato metagenómico del microbioma intestinal humano para la toma de decisiones y la prevención. Las primeras aplicaciones serían para la predicción de cáncer colorrectal. Actualmente, no existen soluciones orientadas a la práctica clínica ni tampoco orientadas a la actividad investigadora. El TRL actual es el 2 y, el que se pretende obtener será un TRL7. Respecto a los riesgos tecnológicos y, al tratarse de un campo nuevo, se espera un uso con aplicaciones limitadas inicialmente.
- g. Sistema interoperativo para pacientes con información derivada de sus datos genómicos:** Permitirá un modelo de sistema de información al paciente totalmente innovador, que ofrece una alternativa validada por el sistema de salud a los test directos al consumidor. El TRL actual es el 2 y, el que se pretende obtener será un TRL6. Respecto a los riesgos tecnológicos, podrían plantearse potenciales dificultades para integrarlo con los resultados de los sistemas de salud.
- h. Entorno de investigación de confianza para el uso secundario de datos médicos en investigación clínica que permita la interoperabilidad con agentes del sistema de investigación y que permita la federación de los estudios:** Permitirá la generación de nuevos biomarcadores y sistemas de soporte a la decisión clínica es un campo innovador y, específicamente, permitirá la oportunidad de aprovechar los sistemas de almacenamiento centralizado de datos médicos disponibles en Andalucía, que acumulan datos de más de 13 millones de pacientes. El TRL actual es el 6 y, el que se pretende obtener será un TRL8. Respecto a los riesgos tecnológicos, podrían plantearse problemas respecto al establecimiento de protocolos seguros para el uso externo de datos que requieran una especial protección. Asimismo, la diversidad de estándares y su poca aplicación pueden limitar las posibilidades de uso federado.

6. Impactos esperados

6.1 Impacto en el servicio sanitario público

La presente propuesta pretende convertir a Andalucía en un referente nacional e internacional en materia de desarrollo de Big Data clínico, unido a aplicaciones de Inteligencia Artificial, los cuales son considerados pilares de la Medicina de Precisión.

Todo ello, mediante el desarrollo de una plataforma integral de uso de datos genómicos en Medicina de Precisión, con un diseño y aplicaciones innovadores hasta el momento. Permitirá, además, un manejo más preciso del paciente, con mayor exactitud en los diagnósticos y



decisiones de tratamientos mucho más eficaces, lo que redundará en una mayor calidad en la atención del paciente y una reducción de los costes de tratamientos y hospitalizaciones.

El uso de los datos genómicos en la práctica clínica rutinaria requiere una estrategia completamente distinta a la actual, en la que se incluyan un alto grado de automatización y un software que permita una trazabilidad integral de los pacientes, sus datos genómicos, sus resultados y las decisiones derivadas, incluyendo sistemas de alertas.

La infraestructura propuesta en este proyecto permitirá la secuenciación genómica de pacientes (biopsias, patógenos infecciosos o ambientales y microbiota intestinal) y utilizará la información obtenida para el diagnóstico, la recomendación de tratamiento y el seguimiento de los pacientes. Además, permitirá a los pacientes interoperar con su información genómica de forma pionera para tomar decisiones informadas preventivas o sobre estilo de vida. Es por ello que uno de los aspectos más innovadores de esta plataforma es la posibilidad de transformar datos genómicos no estructurados y de difícil uso, en datos estructurados para facilitar el uso secundario de esta información, junto con el resto del Big data clínico existente en la Base Poblacional de Salud (BPS), de forma que acelere el descubrimiento de biomarcadores y aumente la precisión de la práctica médica en el sistema de salud. Todo ello, como ya se ha dicho, redundará en una mayor calidad en la atención del paciente y una reducción de los costes de tratamientos y hospitalizaciones, permitiendo, además, una vez validada, su implementación a otros sistemas de salud.

6.2 Impacto socioeconómico en la región

La CPI representa una valiosa oportunidad, no solo para la mejora del sistema sanitario andaluz mediante la aplicación de soluciones tecnológicas, sino también para las empresas, a las cuales ayuda a generar soluciones de innovación para la mejora de la calidad de la atención a la ciudadanía. Esto ha llevado a la Junta de Andalucía a manifestar su compromiso y apuesta por este instrumento de la contratación pública a través de la formulación de la Estrategia de CPI en Andalucía, aprobada por Acuerdo de 6 de febrero de 2018 del Consejo de Gobierno (BOJA número 30, 12 de febrero de 2018), y por la puesta en marcha de sendos programas de financiación de proyectos de CPI a través del Programa Operativo Regional del Marco Europeo (PO FEDER Andalucía) (2014-2020 y 2021-2027).

Se espera que el proyecto genere diversos impactos socioeconómicos en Andalucía, entre los que se pueden destacar los siguientes:

- **GENERACIÓN DE EMPLEO:** El proyecto generará empleo durante la fase de desarrollo, ya que será necesario contratar personal especializado, tanto por parte de la entidad



contratante, como por parte de la entidad adjudicataria, que se espera realice la mayor parte o la totalidad de la I+D en la comunidad autónoma de Andalucía.

- **ATRACCIÓN DE LA INVERSIÓN PRIVADA EN I+D A ANDALUCÍA:** se espera una importante contribución del sector privado, tanto durante las fases de desarrollo, como durante las fases posteriores necesarias, hasta llegar a la comercialización del producto resultante.
- **ATRACCIÓN DE INVERSIÓN PRODUCTIVA A ANDALUCÍA:** tal y como se indicaba, se espera una importante generación de empleo y de creación de instalaciones productivas en Andalucía.
- **GENERACIÓN DE SINERGIAS ENTRE LA ADMINISTRACIÓN, GRANDES EMPRESAS, START-UPS Y CENTROS DE INVESTIGACIÓN EN ANDALUCÍA:** la experiencia indica que en el desarrollo de un producto como el que se persigue en este proyecto implica la participación del servicio público de salud, startups o spin-offs, infraestructuras científicas del Sistema Sanitario Público de Andalucía, y otros centros y departamentos de I+D ajenos al SSPA pero que realicen la inversión en I+D necesaria en nuestra comunidad.

6.3 Impacto medioambiental en la región

La digitalización de los sistemas de salud supone, en general, un impacto positivo a nivel medioambiental. Primordialmente, el ahorro de papel derivado de la digitalización de los procesos supone una de las consecuencias más inmediatas e importantes. Esto, a su vez, tiene una implicación directa en la bajada de las emisiones de CO₂.

Por otro lado, la plataforma que se pretende desarrollar gracias a este proyecto, tiene con finalidad proporcionar un manejo mucho más preciso del paciente, con mayor exactitud en los diagnósticos y decisiones de tratamientos mucho más eficaces.

Todo ello, redundará, en:

- La optimización y reducción de pruebas diagnósticas que se deban realizar a los pacientes. Esto lleva asociado, por un lado, el ahorro de material fungible, lo cual deriva, a su vez, en reducir la demanda de materiales desechables que, en la mayoría de los casos, son contaminantes. Se genera, así, un impacto medioambiental positivo, gracias al ahorro en el suministro de estos materiales, e indirectamente en su fabricación, así como la consecuente contaminación que generan al convertirse en desechos.
- La mejora en los tratamientos de los pacientes, lo cual puede suponer un importante ahorro en medicamentos. Gracias a la Medicina Personalizada, dichos tratamientos están más dirigidos al paciente, evitando en mayor medida el tener que probar diferentes alternativas hasta encontrar la más adecuada. Esto deriva en un impacto medioambiental positivo al disminuir los contaminantes que estos generan, no solo en el proceso de



fabricación, sino en el proceso que siguen una vez son desechados, sus restos o aquellos caducados.

- La reducción de las hospitalizaciones. Al mejorar la calidad de la atención del paciente, los diagnósticos y los tratamientos, los pacientes disminuirán sus desplazamientos a los centros hospitalarios. Esto, igualmente, generará un impacto medioambiental positivo por diferentes razones como la disminución de los traslados al hospital (contaminación por transporte, por ejemplo), o la reducción de los impactos que generan los pacientes ingresados en el hospital, por ejemplo, recursos utilizados, residuos generados, energía consumida en su atención sanitaria, etc.

7. Resultados finales esperados

Los datos genómicos constituyen el paradigma de datos voluminosos y complejos, de manejo complicado, pero con un gran potencial para su uso clínico. Es por ello que la generación de software de manejo de los datos es un campo en constante evolución. El objetivo principal de esta propuesta es convertir a Andalucía en un referente nacional e internacional en materia de desarrollo de Big Data Clínico, unido a aplicaciones de Inteligencia Artificial (IA), los cuales son considerados pilares de la Medicina de Precisión o P4 -Personalizada, Predictiva, Preventiva y Participativa-.

Se espera, por tanto, obtener una plataforma integral de uso de datos genómicos en Medicina de Precisión, la cual permitirá un manejo mucho más preciso del paciente, con mayor exactitud en los diagnósticos y decisiones de tratamientos mucho más eficaces, lo que redundará en una mayor calidad en la atención del paciente y una reducción de los costes de tratamientos y hospitalizaciones.

De manera más específica, los desarrollos o resultados finales esperados son:

- Sistema centralizado de pre-procesamiento de datos genómicos.
- Sistema de manejo de datos genómicos para su uso primario en diagnóstico de enfermedades hereditarias.
- Sistema de manejo de datos genómicos para su uso primario en decisión de tratamiento en cáncer.
- Sistema de manejo de datos genómicos de patógenos virales y bacterianos para el manejo de pacientes infecciosos.
- Sistema de vigilancia epidemiológico basado en datos genómicos de patógenos virales ambientales.
- Sistema de manejo de datos genómicos de la microbiota para la toma de decisiones y la prevención con los pacientes.
- Sistema interoperativo para pacientes con información derivada de sus datos genómicos.



- Entorno de investigación de confianza para el uso secundario de datos médicos en investigación clínica que permita la interoperabilidad con agentes del sistema de investigación y que permita la federación de los estudios.

En definitiva, se espera poder ofrecer una solución a esta necesidad no cubierta que garantice, entre otros:

- 1) Escalabilidad: al ser manejado por el servicio que realiza el consejo genético o da el informe de patología, podría extenderse a todos los hospitales que cuenten con dichos servicios, después de que sus correspondientes especialistas tuviesen la formación adecuada, gracias a la conectividad existente entre ellos.
- 2) Equidad: al ser escalable a los servicios de genética y de patología de todos los hospitales, este nuevo servicio se ofrecería a toda la ciudadanía.
- 3) Sostenibilidad: la escalabilidad de la personalización en la práctica de la medicina, junto con el abaratamiento de los costes de la secuenciación, hacen que el uso de datos genómicos en la práctica clínica sea sostenible e incluso coste-efectivo, como prueban diversos estudios.

Innovación abierta: al tratarse de un campo en continua innovación, surgen constantemente nuevas formas de interpretar los hallazgos en la secuencia genómica, así como de extraer información relevante de ella, por lo que las aplicaciones bioinformáticas usadas para la interpretación de dichos datos están en constante evolución.