

REVISIÓN

DATOS Y ALGORITMOS QUE NOS PERMITEN ENTENDER LA ENFERMEDAD Y AYUDAN A CURARLA

DATA AND ALGORITHMS THAT ALLOW US TO UNDERSTAND DISEASE AND HELP CURE IT

Joaquín Dopazo^{1,2,3}

1. BarcelonaBeta Brain Research Center (BBRC), Barcelona, España.

2. Hospital del Mar Research Institute Barcelona (HMRIB), Barcelona, España

3. CIBER Enfermedades Raras (CIBERER), Instituto de Salud Carlos III, España.

Palabras clave:

Gemelos digitales;
Modelos
mecanísticos;
IA explicable;
Datos del mundo real;
Reposicionamiento
de fármacos;
Medicina de
precisión.

Keywords:

Digital twins;
Mechanistic models;
Explainable AI;
Real-world data;
Drug repurposing;
Precision medicine.

Resumen

La posibilidad de construir modelos virtuales de los pacientes, o gemelos digitales, representa una de las promesas más ambiciosas de la medicina de precisión. Sin embargo, el organismo humano es un sistema de enorme complejidad, todavía insuficientemente conocido para ser simulado de forma completa. Este artículo resume una estrategia intermedia y operativa: combinar el conocimiento acumulado en mapas moleculares con modelos matemáticos mecanísticos, inteligencia artificial explicable y datos clínicos del mundo real. Estos enfoques permiten inferir funciones celulares, identificar mecanismos de enfermedad, priorizar dianas terapéuticas y proponer oportunidades de reposicionamiento farmacológico. A partir de ejemplos en cáncer, COVID-19 y enfermedades raras, se discute cómo los datos pueden transformarse en evidencia y conocimiento accionable.

Abstract

The possibility of building virtual models of patients, or digital twins, is one of the most ambitious promises of precision medicine. However, the human body is an extremely complex system and remains insufficiently understood to be fully simulated. This article summarizes an intermediate and operational strategy: combining existing knowledge encoded in molecular maps with mechanistic mathematical models, explainable artificial intelligence and real-world clinical data. These approaches make it possible to infer cellular functions, identify disease mechanisms, prioritize therapeutic targets and propose opportunities for drug repurposing. Using examples from cancer, COVID-19 and rare diseases, the article discusses how data can be transformed into evidence and actionable knowledge.

INTRODUCCIÓN

La medicina contemporánea se encuentra en un punto de inflexión. Durante décadas, la biomedicina ha acumulado datos moleculares, clínicos, epidemiológicos y de imagen a una escala sin precedentes. Al mismo tiempo, la inteligencia artificial ha demostrado una capacidad creciente para detectar patrones en conjuntos de datos complejos (1). Esta convergencia ha alimentado una idea especialmente atractiva: la posibilidad de construir modelos virtuales de los pacientes que permitan anticipar la evolución de una enfermedad, ensayar intervenciones *in silico* y seleccionar la opción terapéutica más adecuada antes de aplicarla en la práctica clínica (2).

El concepto de gemelo digital procede de la ingeniería y se refiere a una réplica virtual de un objeto o sistema físico que puede ser utilizada para simular su comportamiento en condiciones alternativas. En medicina, esta idea resulta extraordinariamente sugerente, pero también problemática. Un motor, un avión o una infraestructura industrial pueden ser modelados porque conocemos con precisión sus componentes, sus relaciones causales y las leyes físicas que gobiernan su funcionamiento. El organismo humano, por el contrario, es un sistema multiescala, adaptativo, heterogéneo y sólo parcialmente conocido. Aunque se pueden construir simuladores útiles para dominios concretos, como la planificación quirúrgica o el entrenamiento clínico, todavía estamos lejos de simular de forma completa el comportamiento de un paciente real (3)

Autor para la correspondencia

Joaquín Dopazo

BarcelonaBeta Brain Research Center (BBRC)

C/ Wellington, 30 · 08005 Barcelona

E-Mail: jdopazo@barcelonabeta.org

Esta limitación no invalida el valor de los modelos, pero obliga a formular una pregunta más precisa. No se trata de preguntarse si podemos modelar al paciente completo, sino qué partes de la biología humana son suficientemente conocidas para ser modeladas de forma útil. La respuesta se encuentra, al menos en parte, en el conocimiento molecular acumulado durante décadas en rutas de señalización, redes metabólicas y mapas de enfermedad. Estos mapas describen cómo interactúan genes, proteínas y otras moléculas para activar funciones celulares que, en última instancia, contribuyen a fenotipos fisiológicos o patológicos (4–7).

El objetivo de este artículo es presentar una aproximación conceptual y aplicada para transformar datos biológicos y clínicos en conocimiento accionable. La estrategia se basa en combinar tres elementos: mapas moleculares que codifican conocimiento previo, modelos mecanísticos que permiten inferir actividad funcional a partir de datos ómicos, e inteligencia artificial explicable capaz de extender el análisis a regiones del sistema que todavía no están representadas en las rutas conocidas. La utilidad de esta aproximación se ilustra con ejemplos en cáncer, COVID-19 y enfermedades raras, así como con la validación mediante datos clínicos del mundo real.

MATERIAL Y MÉTODOS

Este trabajo se plantea como un artículo de síntesis basado en la conferencia impartida en la Real Academia Nacional de Medicina de España y en una línea de investigación centrada en la integración de modelos mecanísticos, aprendizaje automático y datos clínicos. No se presenta aquí un único estudio experimental, sino un marco metodológico que agrupa resultados publicados, herramientas computacionales y ejemplos de aplicación.

El primer componente del marco es el uso de mapas moleculares. Recursos como KEGG, Reactome, WikiPathways y Disease Maps organizan el conocimiento biológico en forma de rutas, circuitos y relaciones dirigidas entre componentes moleculares (4–7). En el contexto de las rutas de señalización, estas relaciones pueden interpretarse como circuitos que conectan receptores, intermediarios y efectores funcionales. El punto esencial es que el fenotipo celular no deriva de la alteración aislada de un gen, sino de la actividad coordinada de circuitos moleculares completos.

El segundo componente es el modelado mecanístico de la actividad de las rutas, un modelado matemático que incluye las relaciones de causalidad. Algoritmos como HiPathia utilizan la expresión génica como

aproximación a la actividad de proteínas que propagan la señal a través de los circuitos definidos en las rutas biológicas para estimar la actividad de funciones celulares concretas (8–11). De este modo, una medida difícil de interpretar de forma directa, como la expresión de miles de genes, se transforma en una representación funcional más cercana a los procesos biológicos relevantes: proliferación, apoptosis, respuesta inmune, inflamación, reparación del ADN o metabolismo, entre otros.

El tercer componente es la inteligencia artificial explicable. Los modelos matemáticos solo pueden operar sobre la parte del sistema que ha sido previamente descrita en el conocimiento biológico disponible, representado en las rutas biológicas. Sin embargo, gran parte del genoma y muchas relaciones funcionales no están descritos aún y por lo tanto no figuran en ninguna ruta biológica. Para explorar ese espacio, se pueden emplear modelos de aprendizaje automático que “aprendan” las relaciones entre los genes de función y/o actividad desconocidos y los genes ya incluidos en las rutas, añadiéndolos así al catálogo de conocimiento existente. Si esto fuese factible se podría aprender toda la biología de los datos. Lamentablemente, el número de combinaciones posibles de conexiones entre los genes es una cifra tan astronómica, que por muchos datos que se tuviesen sería imposible este aprendizaje. Esto es lo que se llama “la maldición de la dimensionalidad” en IA (12). Sin embargo, se puede reducir la complejidad del problema tratando de “aprender” estas relaciones de una en una en unos pocos casos (13). Es importante el uso de técnicas de explicabilidad, como SHAP, que permiten identificar qué variables contribuyen a una predicción y formular hipótesis biológicas susceptibles de validación experimental o clínica.

Finalmente, el marco incorpora validación con datos del mundo real. Las predicciones derivadas de modelos moleculares e inteligencia artificial deben contrastarse frente a datos clínicos observacionales, idealmente en entornos seguros de tratamiento de datos que minimicen el riesgo de reidentificación y eviten la extracción innecesaria de información sensible. Este principio está alineado con la lógica del Espacio Europeo de Datos Sanitarios y con la necesidad de convertir el uso secundario de datos clínicos en evidencia robusta, auditada y socialmente aceptable (14).

RESULTADOS

El resultado conceptual de esta aproximación es la sustitución de una visión centrada en el gen, heredada de nuestra capacidad de medir propiedades de genes individuales, por una visión

holística, más centrada en la funcionalidad de la célula. En numerosas enfermedades, incluida la oncogénesis, la información relevante no reside únicamente en qué genes están alterados, sino en qué funciones celulares se activan o inhiben como consecuencia de dichas alteraciones. Esta idea conecta con el marco de los *hallmarks* del cáncer, en el que procesos como la proliferación sostenida, la evasión de la apoptosis o la inestabilidad genómica constituyen capacidades funcionales adquiridas por las células tumorales (15). El modelado de rutas permite cuantificar estas funciones y compararlas entre condiciones, pacientes o estadios de enfermedad.

Un ejemplo ilustrativo, descrito en la publicación original de la metodología, procede del cáncer. La actividad de los circuitos relacionados con distintos *hallmarks* de cáncer, como la replicación del ADN o la resistencia a la apoptosis, etc. (ver Figura 1) está asociada significativa-

mente a la supervivencia de los pacientes, como se puede ver en el estudio realizado tomando datos de expresión génica de biopsias de cáncer de riñón disponibles en el repositorio del genoma del cáncer (16). Además, esta actividad está más asociada al pronóstico que la expresión individual de cualquiera de los genes que lo integran. Esta observación refleja una propiedad fundamental de los sistemas biológicos: el sistema es más que la suma de sus partes. Diferentes tumores pueden utilizar programas de expresión génica distintos para activar una misma función celular, lo que hace necesario analizar la arquitectura funcional de las rutas y no solo sus componentes aislados (17).

Más allá de la mera descripción de la biología detrás del cáncer, los modelos matemáticos mecanísticos ofrecen la posibilidad de realizar predicciones causales. Una vez que un circuito molecular está representado como modelo mecanístico, se

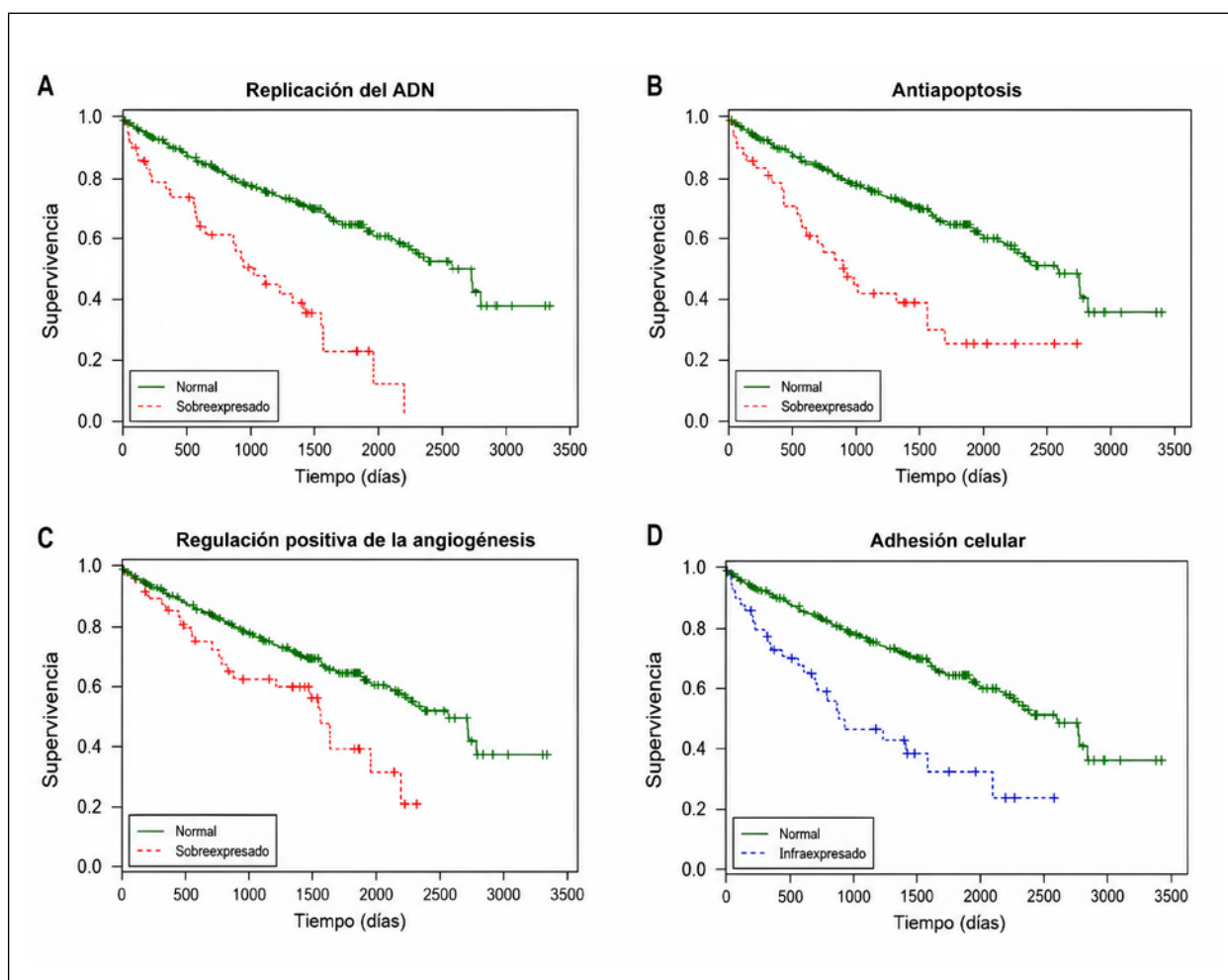


Figura 1. Asociación entre la actividad de funciones celulares y la supervivencia de los pacientes. Curvas de supervivencia de Kaplan-Meier correspondientes a cuatro funciones biológicas relevantes inferidas a partir de circuitos funcionales elementales. En todos los paneles, la línea verde continua representa el grupo con actividad normal, mientras que las líneas discontinuas representan alteraciones funcionales: sobreexpresión en (A) replicación del ADN, (B) antiapoptosis y (C) regulación positiva de la angiogénesis, e infraexpresión en (D) adhesión celular. En los cuatro casos, las alteraciones funcionales se asocian con una menor supervivencia en comparación con el estado normal. Las cruces sobre las curvas indican observaciones censuradas. El tiempo se expresa en días.

pueden simular perturbaciones, como la inhibición o sobreexpresión de genes concretos, y estimar sus consecuencias funcionales. Estas simulaciones no equivalen a un gemelo digital completo del paciente, pero permiten explorar de forma racional intervenciones moleculares dentro de un contexto biológico definido. En estudios previos, este tipo de aproximación ha permitido priorizar módulos funcionales y dianas terapéuticas que posteriormente fueron validadas experimentalmente, demostrando la exactitud de las predicciones (18).

Como se ha comentado, una limitación de este tipo de modelización es que depende del conocimiento previo del sistema descrito en rutas biológicas (de señalización, metabólicas, etc.). La extensión de los modelos mediante aprendizaje automático se presenta en un caso de estudio con la COVID-19. La disponibilidad del *COVID-19 Disease Map* proporcionó un repositorio estructurado de mecanismos moleculares implicados en la infección por SARS-CoV-2 y en sus consecuencias inmunológicas, inflamatorias y de coagulación (19). Sobre esta base, se construyeron modelos mecanísticos capaces de inferir la actividad de circuitos relacionados con la enfermedad a partir de datos de expresión génica (20). Posteriormente, la combinación con inteligencia artificial permitió priorizar fármacos cuyas dianas podían modular funciones relevantes del mapa de enfermedad de la COVID-19 (13).

El resultado más importante procede de la validación de estos modelos con datos clínicos del mundo real. Las predicciones de reposicionamiento farmacológico para COVID-19 fueron contrastadas con una cohorte retrospectiva de pacientes hospitalizados, evaluando la asociación entre tratamientos y supervivencia a 30 días mediante análisis ajustados por covariables clínicas. La observación de un enriquecimiento significativo de fármacos predichos como reposicionables entre aquellos asociados a un efecto protector proporciona un ejemplo de cierre del ciclo entre modelo, predicción y evidencia clínica (21,22). Este tipo de validación no sustituye a los ensayos clínicos, pero ayuda a priorizar hipótesis y a generar evidencia preliminar en contextos donde la experimentación directa es costosa, lenta o limitada.

Otro ámbito importante de aplicación de estos modelos es el de las enfermedades raras, donde la escasez de pacientes, la fragmentación de la información y la dificultad de realizar ensayos convencionales hacen especialmente atractivas las estrategias de reposicionamiento farmacológico. La aplicación sistemática a conjuntos amplios de enfermedades raras permite pasar de una investigación artesanal, centrada en una enfermedad cada vez, a una generación más metódica de hipótesis terapéuticas. Algunos ejemplos de esta aplicación se han hecho y validado en Anemia de Fanconi (23) o en retinitis pigmentosa (24).

Herramientas como drexml integran modelos mecanísticos, aprendizaje automá-

tico y explicabilidad para identificar dianas capaces de modular funciones relevantes de una enfermedad y priorizar fármacos ya conocidos (25).

DISCUSIÓN

Los modelos biomédicos no deben juzgarse por su aspiración a reproducir toda la realidad, sino por su utilidad para responder preguntas concretas. La conocida afirmación de George Box, según la cual todos los modelos son incorrectos, pero algunos son útiles (26), resulta particularmente adecuada en medicina. Un modelo mecanístico de una ruta molecular no pretende capturar la totalidad del organismo, pero puede ser útil para interpretar una alteración genómica, estimar el impacto funcional de una perturbación o priorizar una diana terapéutica. No obstante, existen limitaciones importantes. La primera es la incompletitud del conocimiento biológico. Solo una parte del genoma puede ser actualmente modelada mediante rutas moleculares suficientemente detalladas. Por ello, esta perspectiva permite reconciliar dos tradiciones que a veces se presentan como opuestas. Por un lado, la biología mecanística se basa en hipótesis explícitas, relaciones causales y conocimiento acumulado. Por otro, la inteligencia artificial identifica patrones en grandes volúmenes de datos y puede descubrir asociaciones que no habían sido formuladas previamente. La combinación de ambas aproximaciones es más potente que cualquiera de ellas por separado. El conocimiento previo reduce el espacio de búsqueda y mejora la interpretabilidad; el aprendizaje automático permite explorar regiones del sistema pobremente caracterizadas y proponer nuevas hipótesis.

Otro aspecto a tener en cuenta es la necesidad de validación. Una predicción computacional, incluso si es mecanísticamente plausible y explicable, no debe confundirse con una demostración experimental o clínica. Cuando esta validación requiere datos del mundo real también deben considerarse limitaciones éticas y organizativas. El uso secundario de datos clínicos exige gobernanza, trazabilidad, minimización de riesgos y control de accesos. Los entornos seguros de tratamiento de datos son una respuesta práctica a esta necesidad, porque permiten analizar información sensible sin desplazarla innecesariamente fuera del sistema que la custodia. En este sentido, la transformación de datos clínicos en evidencia no es solo un problema técnico, sino también institucional y social (14).

Este trabajo también propone un cambio de paradigma en la investigación. En la era de los datos y del florecimiento de la inteligencia

artificial, se puede proponer un flujo de trabajo (Figura 2) que ejemplifica una forma alternativa de generación de conocimiento biomédico basada en la integración iterativa de conocimiento previo, datos moleculares y validación por uso secundario de datos clínicos. En primer lugar, los mapas de enfermedad permiten formalizar el conocimiento biológico existente en forma de redes funcionales. A continuación, los métodos de aprendizaje automático pueden explorar grandes volúmenes de datos genómicos para identificar relaciones entre dianas terapéuticas, circuitos moleculares y funciones celulares alteradas. Finalmente, las hipótesis generadas pueden contrastarse mediante datos del mundo real, procedentes de historias clínicas electrónicas y análisis en entornos seguros. De este modo, el proceso conecta el razonamiento mecanístico con la validación poblacional, facilitando la identificación de oportunidades de reposicionamiento terapéutico y cerrando un ciclo de aprendizaje continuo entre biología, datos y práctica clínica.

CONCLUSIONES

La construcción de gemelos digitales completos de pacientes continúa limitada por la enorme complejidad del organismo humano y por las lagunas persistentes en nuestro conocimiento biológico. Sin embargo, esta limitación no impide desarrollar modelos útiles de componentes concretos del sistema. Los mapas moleculares, combinados con modelos mecanísticos, permiten traducir datos ómicos en actividad celular funcional interpretable y asociable a fenotipos. La inteligencia artificial explicable amplía esta capacidad al ayudar a generar nuevas hipótesis sobre dianas y mecanismos de enfermedad. Finalmente, los datos clínicos del mundo real permiten contrastar esas hipótesis en poblaciones reales y cerrar el ciclo entre predicción, evidencia y decisión biomédica. La promesa no reside en sustituir la medicina por simulaciones, sino en convertir los datos en conocimiento accionable para una medicina más preventiva, precisa y racional.

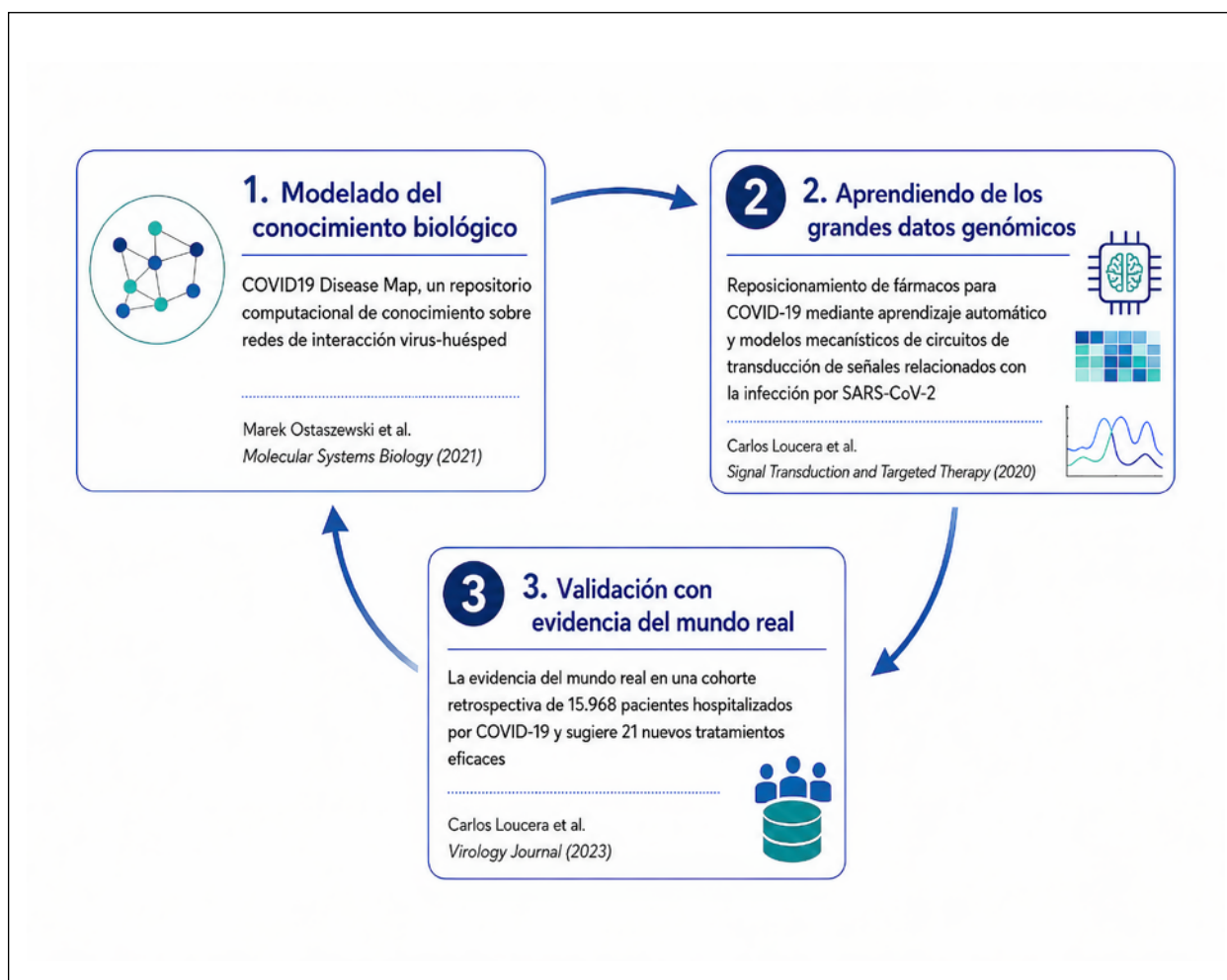


Figura 2. Flujo cíclico de generación y validación de conocimiento biomédico mediante ciencia de datos. El esquema resume una estrategia en tres etapas: modelado del conocimiento biológico disponible, aprendizaje a partir de grandes datos genómicos y validación posterior mediante evidencia del mundo real. El ciclo ilustra cómo los modelos mecanísticos, el aprendizaje automático y los datos clínicos pueden integrarse para generar hipótesis terapéuticas y evaluarlas sin necesidad de experimentos adicionales iniciales.

AGRADECIMIENTOS

El autor agradece a los miembros y colaboradores de las distintas líneas de investigación que han contribuido al desarrollo de los modelos mecanísticos, las herramientas computacionales y los casos de aplicación discutidos en este artículo, así como a las instituciones y agencias financiadoras que han apoyado estos trabajos como el ISCIII (PMP24/00024, IMP/00019, COV20/00788), el Ministerio de Ciencia, Innovación (PID2023-152380OB, TED2021-132346B-I00, PID2020-117979RB-I00), la Consejería de Salud y familias, junta de Andalucía (IE19_259 FPS, P18-RT-3471) y Fundación BBVA (G999088Q).

DECLARACIÓN DE TRANSPARENCIA

El autor/a de este artículo declara no tener ningún tipo de conflicto de intereses respecto a lo expuesto en el presente trabajo.

BIBLIOGRAFÍA

- Altucci L, Badimon L, Balligand JL, et al. Artificial Intelligence and Network Medicine: Path to Precision Medicine. *NEJM AI*. 2025;2(9). <https://doi.org/10.1056/AIRA2401229>.
- Wang H, Arulraj T, Ippolito A, Popel AS. From virtual patients to digital twins in immunoncology: lessons learned from mechanistic quantitative systems pharmacology modeling. *npj Digital Medicine* 2024 7:1. 2024;7(1): 189-. <https://doi.org/10.1038/s41746-024-01188-4>.
- De Domenico M, Allegri L, Caldarelli G, et al. Challenges and opportunities for digital twins in precision medicine from a complex systems perspective. *npj Digital Medicine* 2025 8:1. 2025;8(1): 37-. <https://doi.org/10.1038/s41746-024-01402-3>.
- Kanehisa M, Furumichi M, Sato Y, Kawashima M, Ishiguro-Watanabe M. KEGG for taxonomy-based analysis of pathways and genomes. *Nucleic Acids Research*. 2023;51(D1): D587–D592. <https://doi.org/10.1093/NAR/GKAC963>.
- Ragueneau E, Gong C, Siquin P, et al. The Reactome Knowledgebase 2026. *Nucleic Acids Research*. 2026;54(D1): D673–D681. <https://doi.org/10.1093/NAR/GKAF1223>.
- Ostaszewski M, Niarakis A, Mazei, et al. COVID19 Disease Map, a computational knowledge repository of virus–host interaction mechanisms. *Molecular Systems Biology* 2021 17:10. 2021;17(10): MSB202110387-. <https://doi.org/10.15252/MSB.202110387>.
- Agrawal A, Balci H, Hanspers K, et al. Wiki-Pathways 2024: next generation pathway database. *Nucleic Acids Research*. 2024;52(D1): D679–D689. <https://doi.org/10.1093/NAR/GKAD960>.
- Peña-Chilet M, Esteban-Medina M, Falco MM, et al. Using mechanistic models for the clinical interpretation of complex genomic variation. *Scientific Reports* 2019 9:1. 2019;9(1): 18937-. <https://doi.org/10.1038/s41598-019-55454-7>.
- Rian K, Hidalgo MR, Çubuk C, et al. Genome-scale mechanistic modeling of signaling pathways made easy: A bioconductor/cytoscape/web server framework for the analysis of omic data. *Computational and Structural Biotechnology Journal*. 2021;19: 2968–2978. <https://doi.org/10.1016/J.CSB.2021.05.022>;
- Amadoz A, Hidalgo MR, Çubuk C, Carbonell-Caballero J, Dopazo J. A comparison of mechanistic signaling pathway activity analysis methods. *Briefings in Bioinformatics*. 2019;20(5): 1655–1668. <https://doi.org/10.1093/BIB/BBY040>.
- Hidalgo MR, Cubuk C, Amadoz A, et al. High throughput estimation of functional cell activities reveals disease mechanisms and predicts relevant clinical outcomes. *Oncotarget*. 2016;8(3): 5160–5178. <https://doi.org/10.18632/ONCOTARGET.14107>.
- Crespo Márquez A. The Curse of Dimensionality. *Springer Series in Reliability Engineering*. 2022; 67–86. https://doi.org/10.1007/978-3-030-97660-6_7/SAVE-RESEARCH
- Loucera C, Esteban-Medina M, Rian K, Falco MM, Dopazo J, Peña-Chilet M. Drug repurposing for COVID-19 using machine learning and mechanistic models of signal transduction circuits related to SARS-CoV-2 infection. *Signal Transduction and Targeted Therapy* 2020 5:1. 2020;5(1): 290-. <https://doi.org/10.1038/s41392-020-00417-y>
- Muñoyerro-Muñiz D, Villegas R, de la Oliva V, et al. Ethical and secure evidence generation from regionwide clinical data through a collaborative environment for advancing predictive care. *Frontiers in Public Health*. 2025;13: 1630351. <https://doi.org/10.3389/fpubh.2025.1630351>
- Hanahan D. Hallmarks of Cancer: New Dimensions. *Cancer Discovery*. 2022;12(1): 31–46. <https://doi.org/10.1158/2159-8290.CD-21-1059>
- Strausberg RL, Buetow KH, Emmert-Buck MR, Klausner RD. The Cancer Genome Anatomy Project: Building an annotated gene index. *Trends in Genetics*. 2000;16(3): 103–106. [https://doi.org/10.1016/S0168-9525\(99\)01937-X](https://doi.org/10.1016/S0168-9525(99)01937-X).
- Hidalgo MR, Amadoz A, Çubuk C, Carbonell-Caballero J, Dopazo J. Models of cell signaling uncover molecular mechanisms of high-risk neuroblastoma and predict disease outcome. *Biology Direct* 2018 13:1. 2018;13(1): 16-. <https://doi.org/10.1186/S13062-018-0219-4>
- Cubuk C, Hidalgo MR, Amadoz A, et al. Gene expression integration into pathway modules reveals a pan-cancer metabolic landscape. *Cancer Research*. 2018;78(21): 6059–6072. <https://doi.org/10.1158/0008-5472.CAN-17-2705/653103/AM/GENE-EXPRESSION-INTEGRATION-INTO-PATHWAY-MODULES>

19. Ostaszewski M, Niarakis A, Mazein A, et al. COVID19 Disease Map, a computational knowledge repository of virus-host interaction mechanisms. *Molecular systems biology*. 2021;17(10). <https://doi.org/10.15252/MSB.202110387>
20. Rian K, Esteban-Medina M, Hidalgo MR, et al. Mechanistic modeling of the SARS-CoV-2 disease map. *BioData Mining*. 2021;14(1). <https://doi.org/10.1186/s13040-021-00234-1>
21. Loucera C, Peña-Chilet M, Esteban-Medina M, et al. Real world evidence of calcifediol or vitamin D prescription and mortality rate of COVID-19 in a retrospective cohort of hospitalized Andalusian patients. *Scientific Reports* 2021 11:1. 2021;11(1): 23380-. <https://doi.org/10.1038/s41598-021-02701-5>
22. Loucera C, Carmona R, Esteban-Medina M, et al. Real-world evidence with a retrospective cohort of 15,968 COVID-19 hospitalized patients suggests 21 new effective treatments. *Virology Journal* 2023 20:1. 2023;20(1): 226-. <https://doi.org/10.1186/S12985-023-02195-9>
23. Esteban M, Peña-Chilet M, Loucera C, Dopazo J. Exploring the druggable space around the Fanconi anemia pathway using machine learning and mechanistic models. *bioRxiv*. 2019. <https://doi.org/10.1101/647735>.
24. Esteban-Medina M, Loucera C, Rian K, et al. The mechanistic functional landscape of retinitis pigmentosa: a machine learning-driven approach to therapeutic target discovery. *Journal of Translational Medicine* 2024 22:1. 2024;22(1): 139-. <https://doi.org/10.1186/S12967-024-04911-7>
25. Esteban-Medina M, de la Oliva Roque VM, Herráiz-Gil S, Peña-Chilet M, Dopazo J, Loucera C. drexml: A command line tool and Python package for drug repurposing. *Computational and Structural Biotechnology Journal*. 2024;23: 1129–1143
26. Box GEP. Science and Statistics. *Journal of the American Statistical Association*. 1976;71(356): 791–799. <https://doi.org/10.1080/01621459.1976.10480949>

Si desea citar nuestro artículo:

Dopazo J. Datos y algoritmos que nos permiten entender la enfermedad y ayudan a curarla. *An RANM*. 2026;143(02): 167–173. DOI: 10.32440/ar.2026.143.02.rev04
