Módulo 1

1.2. La Bioinformática. Aplicaciones en Bio-Ciencias y Bio-Salud

*Por **Coral del Val Muñoz***
Profesora Titular de la Universidad de Granada. Departamento de Ciencias de la Computación e Inteligencia Artificial (DECSAI).
Instituto Andaluz Interuniversitario en Data Science and Computational Intelligence (DasCI)

*Por **Pedro Carmona Sáez***
Profesor Ayudante Doctor de la Universidad de Granada. Departamento de Estadística e Investigación Operativa

---

1. INTRODUCCIÓN

Actualmente, podemos encontrar aplicaciones de la Bioinformática en todas las ciencias de la vida y la salud en las que el desarrollo de nuevas tecnologías experimentales, han generado cantidades masivas de datos. Los campos de aplicación donde la bioinformática está teniendo un papel cada vez más importante son muy amplios, como ejemplo:

- Desarrollo y descubrimiento de fármacos
- Farmacogenómica
- Medicina de precisión o personalizada
- Microbiología
- Biocombustibles
- Minería de textos biomédicos
- Tecnología de los alimentos y nutrición
- Desarrollo de bases de datos en biología
- Biología sintética y de sistemas
- Procesamiento de imágenes de imágenes médicas y diagnóstico

En el video de introducción proponemos, a modo ilustrativo, algunas aplicaciones en estas áreas.

2. MEDICINA DE PRECISIÓN O PERSONALIZADA

Si buscamos un campo en el que a la bioinformática se le haya encontrado una gran aplicación es la medicina de precisión o medicina personalizada. La medicina personalizada se resume en la expresión inglesa de "serve the right patient with the right drug at the right time", es decir, administrar el tratamiento correcto al paciente correcto en el momento correcto. En contraposición a la medicina tradicional, donde el diagnóstico se realiza principalmente en base a manifestaciones clínicas de la patología, la medicina personalizada se basa en comprender los mecanismos moleculares que dan origen la enfermedad y establecer el diagnóstico y el tratamiento en base a los mismos.

El campo donde este tipo de aproximación está teniendo un mayor impacto es el cáncer, donde un mismo tumor puede estar originado por diversas alteraciones moleculares o mutaciones, que pueden ser completamente diferentes en pacientes con el mismo tipo tumoral, por lo que un tratamiento puede tener efectos totalmente diferentes en ellos. La caracterización de estas alteraciones a nivel individual permitirá administrar tratamientos personalizados con el consecuente beneficio, tanto médico como económico ya que se evita administrar terapias que serán infectivas.

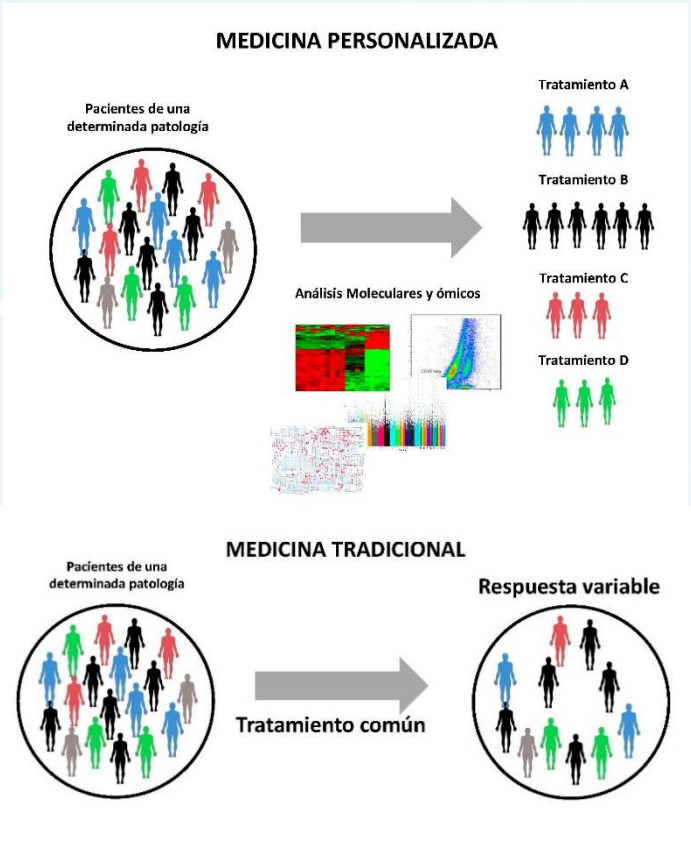

Figura 1. Concepto de medicina de precisión

3. FARMACOGENÓMICA

La farmacogenómica está siendo una disciplina importante para la implementación de la medicina personalizada impulsada por la aplicación de técnicas bioinformáticas y de análisis de datos. La farmacogenómica proporciona información relacionada con la respuesta de un individuo a un medicamento en función de su genética. Se basa en el estudio del genoma de pacientes para establecer terapias más eficaces, predecir respuesta a fármacos o efectos secundarios. En este contexto, se suelen estudiar polimorfismos o variantes que los diferentes individuos tienen en genes que codifican enzimas responsables de la metabolización de fármacos o que son dianas de los mismos para establecer la respuesta al tratamiento o efecto adversos que puedan tener.

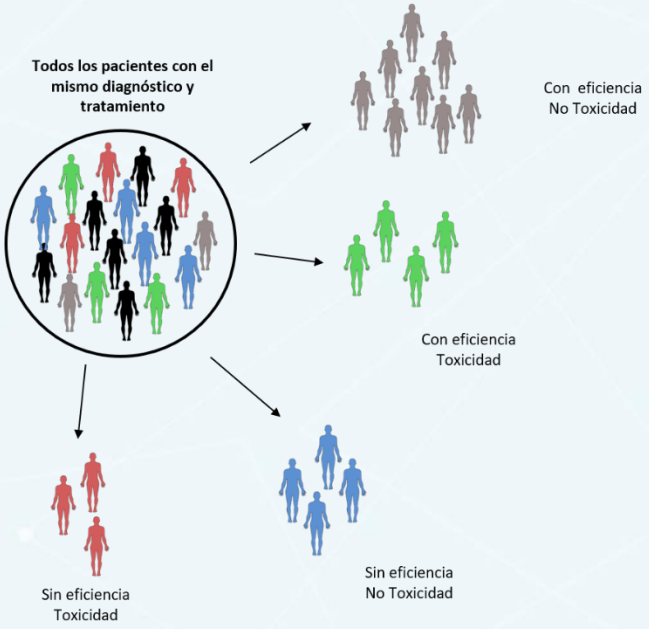

Figura 2. Concepto de farmacogenómica. Un fármaco puede tener diferentes efectos según las alteraciones moleculares por lo que agrupar pacientes en función de las mismas ayudará a tener unas terapias más efectivas

4. DESARROLLO Y DESCUBRIMIENTO DE FÁRMACOS

El descubrimiento y el desarrollo de fármacos es un proceso complejo, que requiere de una alta inversión por parte de las compañías farmaceúticas, en coste y tiempo. La estrategia principal para el desarrollo de medicamentos durante las últimas décadas se ha basado en el detección mediante técnicas de *screening* masivo de la actividad de miles de moléculas simultáneamente para identificar compuestos que muestran actividad. Cuando se prueban más medicamentos, es más probable que

se encuentren resultados, pero estos enfoques están vinculados a enormes costos y baja eficiencia, es decir, una muy baja relación eficiencia / costo.

La información previa sobre los sistemas biológicos y la explotación de datos mediante técnicas bioinformáticas ofrece enormes posibilidades para abaratar y agilizar los procesos de descubrimiento de fármacos. La bioinformática ofrece soluciones en diferentes pasos del este proceso, como:

4.1. ANÁLISIS ESTRUCTURALES Y BIOFÍSICOS DE PROTEÍNAS

Para búsqueda automática en base de datos para la identificación de potenciales candidatos de fármacos que interaccionan con la proteína de interés.

4.2. REPOSICIONAMIENTO DE FÁRMACOS

Se basa en encontrar nuevas aplicaciones para fármacos conocidos. En este sentido, se han usado técnicas bioinformáticas para búsqueda de fármacos que revierten perfiles de expresión génica analizando la información disponible en bases de datos como connectivity map que contiene decenas de miles de firmas de expresión de diferentes compuestos y fármacos (https://clue.io/cmap). Esta área está ahora de máxima actualidad por el potencial de este tipo de aproximaciones para búsqueda de fármacos en el tratamiento de la COVID-19, donde se están encontrando compuestos efectivos que se han venido usando en otras patologías.

4.3. IDENTIFICACIÓN DE DIANAS TERAPÉUTICAS

Uno de los principales objetivos en el análisis de datos ómicos es la identificación de biomarcadores (genes, proteína, metabolitos, etc) que están alterados en una condición patológica. Estos elementos pueden servir de biomarcadores de diagnóstico, pero también son potenciales dianas terapéuticas que pueden ser explorados para el desarrollo de terapias específicas.

5. MICROBIOLOGÍA Y METAGENÓMICA: DE LAS BIO-BATERÍAS A LA BIORREMEDIACIÓN

Actualmente, se conoce la secuencia de más de 250.000 genomas microbianos, y están en marcha multitud de grandes proyectos de investigación en metagenómica. Dada las implicaciones de los microbios en salud, energía, medio ambiente y en las aplicaciones industriales, el estudio de su material genético permitirá comprender estos microbios a un nivel muy fundamental. Estos conocimientos son el primer pilar para aislar los genes que les otorgan sus capacidades únicas para sobrevivir en condiciones extremas o realizar procesos de interés. Dos ejemplos son:

5.1. NUEVAS FUENTES DE ENERGÍA

Es uno de los campos donde diversos proyectos se están desarrollando para entender y mejorar el metabolismo de microorganismos productores de corriente mediante el uso de la bioinformática. Así por ejemplo se han identificado microbios capaces de generar energía a partir de la luz como Chlorobium tepidum o mediante procesos de reducción de iones metálicos a través de biopelículas metabólicamente activas como en el caso de *Geobacter sulfurreducens.* Los nuevos avances y la comprensión de cómo tiene lugar el transporte largo de electrones entre bacterias nos acerca cada vez más la creación de bio-baterías que puedan aumentar la producción de energía verde.

Figura 3. El estudio bioinformático del material genético y las rutas metabólicas de determinados microbios abre la posibilidad a la creación de pilas naturales que puedan aprovechar aguas residuales, o lodos para la generación de energía.

5.2. BIORREMEDIACIÓN

La biorremediación explora el potencial microbiano para la biodegradación de compuestos xenobióticos y residuos, ya que estos son capaces de degradar contaminantes y por tanto restaurar de manera eficiente y efectiva las condiciones ambientales originales. La bioinformática se ha utilizado en este campo para determinar estructuras y rutas de biodegradación de compuestos xenobióticos. Un ejemplo es el caso de *Deinococcus radiodurans* es la bacteria más resistente, este organismo se está estudiando por su posible potencial en la limpieza de sitios contaminados con radiación y/o químicos tóxicos. Otro ejemplo es el uso de *E. coli* y *Pseudomonas aeruginosa* para la degradación de los derivados del petróleo.

6. BIOLOGÍA SINTÉTICA: DESDE LA BIOMEDICINA A LOS BIOCOMBUSTIBLES

La biología sintética es una disciplina emergente que tiene como objetivo diseñar nuevos sistemas biológicos o rediseñar los existentes con determinados fines. Es una disciplina interdisciplinar que integra ciencias físicas, químicas, computacionales y biomédicas. Aparte de integrar técnicas bioinformáticas, esta disciplina tiene su germen en la mejor comprensión de los sistemas biológicos y genomas gracias a los análisis de los mismos que se han venido haciendo estos años mediante generación de conocimiento a partir del análisis de grandes cantidades de datos.
En los últimos años, la biología sintética está experimentando un considerable avance en diversas áreas, algunos ejemplo son:

6.1. BIOMEDICINA

Se han diseñado organismos con algunas capacidades interesantes, un ejemplo son los circuitos sintéticos que hacen que un género de bacteria, *Yersinia*, adquiera la capacidad de invadir células tumorales. Estas bacterias se han programado con éxito para desencadenar la expresión de un fármaco para reprimir el crecimiento tumoral en ratones.

6.2. PRODUCCIÓN DE SUSTANCIAS DE INTERÉS

Los microorganismos se han utilizado para durante décadas para ello, y el avance de la biología sintética proporciona nuevas herramientas y estrategias para mejorar la eficiencia y las capacidades de estos organismos. Por ejemplo, se han diseñado levaduras con nuevas vías metabólicas capaces de sintetizar un compuesto determinado mediante ingeniería metabólica e introducción de genes de otros microorganismos.

6.3. FITOMEJORA A TRAVÉS DE LA SELECCIÓN DE MICROORGANISMOS ADAPTADOS A ESTRESES

El cambio climático constituye una gran amenaza para la producción agrícola, afectando a la producción mundial con una mayor repercusión en los países más pobres. La bioinformática está jugando un papel clave en la integración de datos genómicos, proteómicos, lipidómicos, metabolómicos resultantes de la interacción planta-microorganismos para la selección de fenotipos resistentes a estreses abióticos, como el aumento de la temperatura y las sequías, y/o estreses bióticos como la resistencia a hongos, o insectos que permitan mantener la producción agrícola y la diversidad del suelo.

6.4. BIOCOMBUSTIBLES

Es uno de los campos donde la biología sintética y la bioinformática pueden ayudar a obtener grandes avances, y diversos proyectos se están desarrollando para mejorar el metabolismo de microorganismos principalmente microalgas, para producción de hidrógeno o etanol como fuentes de energía, la transformación de residuos en energía o la conversión de energía solar en hidrógeno.

Figura 6. La biología sintética tiene aplicaciones fundamentales en un futuro próximo como son la producción de biocombustibles y cultivos más resistentes y eficientes

La comprensión de la estructura y organización del genoma de los diferentes organismos gracias a técnicas de bioinformática, está permitiendo incluso adentrarse en la síntesis de organismos completos y comprender la base de la vida a un nivel antes inimaginable. En este contexto, en 2019 se ha publicado un artículo en la revista Nature donde un grupo de investigadores crearon una variante de *Escherichia coli* con un genoma sintético de cuatro megabases.

7. TECNOLOGÍA DE ALIMENTOS Y CIENCIAS DEL DEPORTE

Para la caracterización y estandarización de las materias primas, el desarrollo de procesos y la detección de variaciones de lote a lote y el control de calidad del producto final se está extendiendo el uso de proteómica. En el caso de la seguridad biológica y microbiana y el uso de alimentos genéticamente modificados, se está usando la genómica, como muestra por ejemplo la aplicación de estas técnicas para la localización del punto cero de la infección en el brote de *Listeria monocytes* en el 2019 en Andalucía.

En el contexto de nutrigenómica la bioinformática también juega un papel fundamental. La nutrigenética se puede definir como estudios de nutrición y herencia, mientras que la nutrigenómica es el estudio de las interacciones mutuas entre las moléculas, los genes y la función genética de la dieta. La importancia de la bioinformática en esta área viene determinada por la gran acumulación de datos disponibles, y existen numerosas bases de datos y paquetes de software bioinformático dedicados en exclusiva en este contexto. En este contexto, destaca también el estudio de enfermedades complejas como por ejemplo la obesidad. Se están estudiando interacciones entre el patrón de la dieta y los factores genéticos (variaciones genéticas) que condicionan la forma que las personas responden a la dieta con el fin de prevenir o reducir la obesidad o mejorar su rendimiento deportivo.
También se han realizado muchos estudios en la mejora nutricional de los productos derivados de animales o plantas, como la transferencia genética en arroz para incrementar niveles de vitamina A, hierro y otros micronutrientes.

Figura 7. Entender las interacciones entre genética y nutrición, permitirán prevenir patologías y entender mejor las características individuales del efecto de la dieta y patrones de alimentación

8. PROCESAMIENTO DE IMÁGENES MÉDICAS

Los datos de imágenes médicas son una de las fuentes más ricas de información sobre pacientes (e.g. rayos X, tomografías computarizadas, resonancias magnéticas, PET, etc.) y, a menudo, una de las más complejas siendo un desafío su interpretación incluso para el profesional clínico más experimentado. Por ello hay un creciente interés en transformar las imágenes médicas en datos útiles que se puedan utilizar para mejorar la toma de decisiones clínicas.

El aprendizaje automático (machine learning) y la inteligencia artificial implementadas en muchas herramientas bioinformáticas de análisis de imágenes han encontrado un amplio nicho en el diagnóstico biomédico, como por ejemplo la mejora del diagnóstico de cáncer (e.g. cáncer de piel Esteva, 2017), de enfermedades de retina o, en un contexto de máxima actualidad, se están desarrollando algoritmos basados en *deep learning* para el diagnóstico de la COVID-19 mediante el análisis de radiografía de pulmón.
Ejemplos concretos de este tipo de aplicaciones actuales están:

8.1. DETECCIÓN DE CÁNCERES COMUNES MEDIANTE EL ANÁLISIS DE IMÁGENES

Las imágenes médicas a menudo se usan en exámenes preventivos de rutina para cánceres, como el cáncer de mama, de piel y el de colon. Una de las más usadas es la ***tomografía digital de mama (DBT)*** también conocida como mamografía 3D, su uso combinado con MRI y ultrasonido ha mejorado la precisión de los diagnósticos de cáncer de mama minimizando el error y ayudando en la detección temprana del cáncer.

8.2. DETECCIÓN DEL CÁNCER DE PIEL

El melanoma (la forma más mortal de cáncer de piel) es altamente curable si se diagnostica temprano y se trata adecuadamente. Las tasas de supervivencia llegando al 65, si se diagnostica en estadios iniciales. El tratamiento adecuado puede incluso producir una tasa de supervivencia a 5 años de más del 98 por ciento. El algoritmo DL, es uno de los más prometedores corto plazo siendo capaz de detectar en imágenes melanoma con mayor precisión que la mayoría de los expertos.

8.3. ANÁLISIS DE IMÁGENES CARDIOVASCULARES

Su uso ha minimizado la utilización de otras pruebas diagnósticas invasivas como angiografías y cateterismos. La posibilidad de medir en imágenes las diversas estructuras del corazón puede

revelar el riesgo de enfermedades cardiovasculares de un individuo o identificar problemas que deben abordarse mediante cirugía o tratamiento farmacológico.

8.4. DIAGNÓSTICO DE ENFERMEDADES NEUROLÓGICAS

Existen muchas enfermedades neurológicas degenerativas, que presentan afecciones neurológicas similares (e.g. la esclerosis lateral amiotrófica (ELA) y la esclerosis lateral primaria (ELP)) aunque unas son mucho más devastadoras que otras. El poder dar al paciente diagnósticos precisos podría ayudar a evitar diagnósticos erróneos innecesarios, y planificar la atención a largo plazo o los deseos al final de la vida. Los algoritmos en los que se está trabajando intentan simplificar el proceso de catalogación y anotación de las imágenes resaltando resultados sospechosos y índices de riesgo de que las imágenes contengan evidencia de ALS o PLS.

Figura 8. Nuevas metodologías de aprendizaje automático están teniendo resultados muy prometedores en diagnóstico a partir de imagen médica

REFERENCIAS BIBLIOGRÁFICAS

- **Ayers, D., and Day, P.J. (2015).** Systems Medicine: The Application of Systems Biology Approaches for Modern Medical Research and Drug Development (Hindawi).
- **Baumann, N. (2016).** How to use the medical subject headings (MeSH). Int. J. Clin. Pract. 70, 171–174.
- **El Karoui, M., Hoyos-Flight, M., and Fletcher, L. (2019).** Future Trends in Synthetic Biology—A Report. Front. Bioeng. Biotechnol. 7.
- **Katara, P. (2013).** Role of bioinformatics and pharmacogenomics in drug discovery and development process. Netw. Model. Anal. Health Inform. Bioinforma. 2, 225–230.
- **Krallinger, M., Erhardt, R., and Valencia, A. (2005).** Text-mining approaches in molecular biology and biomedicine. Drug Discov. Today 10, 439–445.
- **Malkaram, S.A., Hassan, Y.I., and Zempleni, J. (2012).** Online Tools for Bioinformatics Analyses in Nutrition Sciences. Adv. Nutr. 3, 654–665.
- **Rigden, D.J., and Fernández, X.M. (2020).** The 27th annual Nucleic Acids Research database issue and molecular biology database collection. Nucleic Acids Res. 48, D1–D8.
- **Subramanian, A., Narayan, R., Corsello, S.M., Peck, D.D., Natoli, T.E., Lu, X., Gould, J., Davis, J.F., Tubelli, A.A., Asiedu, J.K., et al. (2017).** A Next Generation Connectivity Map: L1000 Platform and the First 1,000,000 Profiles. Cell 171, 1437-1452.e17.
- **Xia, X. (2017).** Bioinformatics and Drug Discovery. Curr. Top. Med. Chem. 17, 1709–1726.
- **Gillies R, Kinahan P, Hricak H. Radiomics:** images are more than pictures, they are data. Radiology 2016; 278: 563–77.