

2023

Biología de sistemas: introducción



Eduardo Ochoa Hernández
Gladys Juárez Cisneros
Héctor Javier Anselmo Villegas Moreno
Nicolás Zamudio Hernández
Berenice Yahuaca Juárez
Lizbeth Guadalupe Villalon Magallan
Rogelio Ochoa Barragán
Salomón Eduardo Borjas García
...





Biología de sistemas: introducción

Autores:

Eduardo Ochoa Hernández
Gladys Juárez Cisneros
Héctor Javier Anselmo Villegas Moreno
Nicolás Zamudio Hernández
Lizbeth Guadalupe Villalon Magallan
Rogelio Ochoa Barragán
Salomón Eduardo Borjas García
Juan Alejandro Cortez Rangel
Daniela Fernández Gómez
Marco Antonio Alemán Méndez
Nestor Alejandro Muñoz Ruiz
Neftali Rangel García
Nicolás Abraham Zamudio Durán
Fatima Ariana Cruz Cendejas
Miriam Rodríguez Caratachea
José Fabián Sánchez López
Mitzi Arismel Pérez Díaz
Estrada López Brittany Dayan
Pedro Gallegos Facio
Gerardo Sánchez Fernández

ISBN: 978-607-8416-24-0

Morelia. Michoacán. Enero de 2023



Universidad Michoacana de San Nicolás de Hidalgo

Libertad Académica Nicolaita LAN/QFB

PRESENTA:

Biología de sistemas: introducción

Autores:

Eduardo Ochoa Hernández
Gladys Juárez Cisneros
Héctor Javier Anselmo Villegas Moreno
Nicolás Zamudio Hernández
Lizbeth Guadalupe Villalon Magallan
Rogelio Ochoa Barragán
Salomón Eduardo Borjas García
Juan Alejandro Cortez Rangel
Daniela Fernández Gómez
Marco Antonio Alemán Méndez
Nestor Alejandro Muñoz Ruiz
Neftali Rangel García
Nicolás Abraham Zamudio Durán
Fatima Ariana Cruz Cendejas
Miriam Rodríguez Caratachea
José Fabián Sánchez López
Mitzi Arismel Pérez Díaz
Estrada López Brittany Dayan
Pedro Gallegos Facio
Gerardo Sánchez Fernández

Título original de la obra:

Ochoa H. E., *et al.* (2023). **Biología de sistemas: introducción**. Morelia: UMSNH-LAN

Copyright © 2023

Tzintzuntán No. 173 Col. Matamoros C.P. 58240, Edificio E planta alta Morelia, Michoacán. México. MX

Teléfono 443 2 731569. Email: ehqfb@yahoo.com.mx

ISBN: 978-607-8416-24-0



Programa: Profesor escritor.

Esta obra fue publicada originalmente en Internet bajo la categoría de contenido abierto sobre la URL: <http://www.libertadacademica.com> mismo título y versión de contenido digital. Este es un trabajo de autoría publicado sobre Internet Copyright © 2023 por la LAN/UMSNH protegido por las leyes de derechos de propiedad de los Estados Unidos Mexicanos. No puede ser reproducido, copiado, publicado, prestado a otras personas o entidades sin el permiso explícito por escrito de LAN o por los Autores.



Directorio

Dra. Yarabí Ávila González
Rectora

Dr. Zoe Tamar Infante Jiménez
Secretario General

Dra. Angélica Guadalupe Zamudio de la Cruz
Secretaria Académica

D.C.E. Javier Cervantes Rodríguez
Secretario Administrativo

Mtra. Mónica Gutiérrez Legorreta
Secretaria Auxiliar

C.P. Enrique Eduardo Román García
Tesorero

Dr. Miguel Ángel Villa
Secretario de Difusión Cultural y Extensión Universitaria

Dr. Raúl Carrera Castillo
Abogado General

Mtra. María Etelvina Rubio Rangel
Contralora

Dr. Jesús Campos García
Coordinador de la Investigación Científica

Dr. Juan José Osorio Ramos
Coordinador de Bachillerato

Dra. Karina Mariela Figueroa Mora
Directora de Tecnologías de la Información y Comunicación

M.I. Cindy Lara Gómez
Coordinadora de Planeación, Infraestructura y Fortalecimiento Universitario

Dra. Arauci Siloe López Huéramo
Secretaria de Acceso a la información

Contenido

Módulo 0. El universitario y sus demonios	1
0. Los demonios que desafían la explicación racional científica	1
0.1 La biología en las nanomáquinas celulares	7
0.2 A pasos de mayor complejidad	10
0.3 La predictibilidad y lo ontológico como generadores de tecnología	13
0.4 Darse la mano empiristas y teóricos	17
0.5 Ninguna explicación científica es para siempre	21
0.6 Modelado matemático de sistemas biológicos	23
0.7 Punto de partida	27
0.8 Introducción a los mecanismos	28
0.8.1 Mecanismos	30
0.8.2 Estado óptico de los mecanismos	32
0.8.3 Actividades y causas	35
0.8.4 Actividades y leyes	36
0.8.5 Actividades y leyes, jerarquías en bocetos	37
0.8.6 Tocando fondo	37
0.9 Los sistemas biológicos son altamente autónomos	39
0.9.1 Fundamento químico de la vida	43
0.9.2 Fundamento físico de la vida	46
0.9.3 Código: la tercera dimensión de la ley de la naturaleza	49
0.9.4 Relatividad de código, energía y masa	52
Módulo 1. Introducción a la biología de sistemas	58
1.1 ¿Qué es la biología de sistemas?	58
1.2 Temas de filosofía de la Biología de Sistemas	63
1.2.1 Reduccionismo y lo emergente	63
1.2.2 Modelado matemático y computacional en biología de sistemas	65
1.2.3 Explicaciones y principios de diseño	68
1.2.4 Biología de sistemas funcional y evolutiva	70
1.2.5 Colaboración interdisciplinaria y educación científica	72
1.2.6 Modos de investigación y estructura institucional	74
1.2.7 ¿La biología de sistemas como el futuro de la medicina?	75
1.3 Perspectivas futuras de la filosofía de la biología de sistemas	77
Módulo 2. La biología de sistemas vista panorámica	80
2.1 Contexto	80
2.1.2 Objetivos y conceptos	82

2.1.3 Estrategias	84
2.2 Introducción a la genómica funcional, proteómica, metabolómica y bioinformática	86
2.2.1 Tecnologías de secuenciación	86
2.2.2 Espectrometría de masas	87
2.2.3 Bioinformática	89
2.3 Genómica funcional, proteómica y metabolómica	91
2.3.1 Epigenómica	91
2.3.2 Transcriptómica	94
2.3.3 Proteómica	96
2.3.4 Metabolómica y lipidómica	99
2.3.4.1 Metabolómica	99
2.3.4.2 Lipidómica	100
2.4 Métodos y herramientas	102
2.4.1 Agrupación	103
2.4.2 Selección de funciones	104
2.4.3 Análisis de predicción	106
2.4.4 Construyendo redes y caminos a partir de la literatura	107
2.4.5 Análisis de ruta	109
2.5 Conclusiones	110
Referencias	111

El creciente interés en los sistemas por parte de los informáticos y los defensores de la teoría general de sistemas a mediados del siglo XX comenzó a proporcionar herramientas conceptuales que podrían abordar sistemas completos, pero estos siguieron siendo, en su mayor parte, pequeños pasos que a menudo perdían contacto con la biología real. El surgimiento de la **biología de sistemas** en el siglo XXI refleja el mayor desarrollo de herramientas de investigación que pueden brindar a los integrantes una agenda de investigación que realmente pueda complementar la empresa altamente exitosa del mecanicista de identificar los componentes y caracterizar sus operaciones. Hay dos tipos distintos de metodologías de investigación que ha presentado la biología de sistemas. La primera es la capacidad de recopilar y analizar grandes cantidades de datos de modo que se pueda obtener información sobre, por ejemplo, el patrón de expresión de un gran número de genes o actividades de un gran número de moléculas en las células. La segunda, es el desarrollo de herramientas matemáticas que permitan a los investigadores representar la organización y el comportamiento de sistemas de gran número de componentes que interactúan de forma no lineal y están organizados de forma no secuencial. Estos incluyen las herramientas de teoría de grafos, modelado computacional y teoría de sistemas dinámicos. En este sentido, el discurso asociado a la biología de sistemas ha superado los resultados obtenidos hasta la fecha. Esto ha despertado el escepticismo entre los críticos. En muchos aspectos, esta historia temprana de la biología de sistemas se parece a la de la inteligencia artificial (IA). La IA ha logrado un progreso muy sustancial y, de hecho, muchas de sus herramientas se están utilizando en biología de sistemas. Sin embargo, no ofrece soluciones instantáneas. Pero lo emocionante es que ha proporcionado nuevas herramientas de investigación cuyo potencial requerirá tiempo para realizarse.

Lo que más nos interesa es cómo la biología de sistemas puede revelar prácticas científicas que no se han caracterizado en la biología que le antecede. Tenemos la esperanza de que la matemática pueda hacer contribuciones a la ciencia real, pero dado el sórdido historial de aplicación del análisis dentro de la ciencia, creemos que los biólogos deberían proceder con gran humildad y cautela. Un problema que aflige a cualquier esfuerzo interdisciplinario es que las personas en una disciplina adoptan lo que se propone en otra disciplina como decisivo cuando en realidad es uno de los muchos puntos de vista y, a menudo, representa una etapa en el desarrollo de esa disciplina. El remedio es comprometerse con la literatura original desde una

perspectiva de modelado dentro de la revolución científica de la causalidad.

El tema más olvidado, en gran parte debido a los desafíos epistémicos para comprenderlo, es el **papel de la organización de los sistemas** en dar lugar a su comportamiento dinámico. En los últimos años del siglo XX, los filósofos de la biología comenzaron a ponerse al día con los dominios de la biología que habían estado buscando explicaciones mecanicistas descomponiendo los mecanismos en sus partes componentes y operaciones. En campos como la biología celular y molecular, los investigadores del siglo XX desarrollaron técnicas poderosas para identificar las estructuras componentes de los sistemas biológicos y determinar las operaciones que realizan. Aunque todas las explicaciones de la explicación mecanicista incluían referencias a la organización de partes y operaciones como algo crucial para la capacidad de los mecanismos para generar el fenómeno que los investigadores están tratando de explicar, hubo poca discusión sobre los modos de organización y las consecuencias que los diferentes modos de organización podrían tener para el funcionamiento dinámico del mecanismo. El **desafío epistémico es identificar patrones de organización y determinar sus efectos**. En gran parte como consecuencia del hecho de que los seres humanos desarrollan nuevos pensamientos de forma secuencial, cuando nos acercamos a las interacciones de múltiples componentes, pensamos en ellos llevando a cabo sus operaciones de forma secuencial. Cuando los biólogos representan un mecanismo biológico, a menudo se basan en diagramas de nodos y flechas. Inferir el comportamiento del mecanismo implica seguir un camino a través del diagrama y simular mentalmente los efectos de cada operación. Esta estrategia funciona razonablemente bien cuando la organización es secuencial y las interacciones son lineales. Incluso puede extenderse a casos un poco menos secuenciales, como aquellos en los que una operación se retroalimenta sobre una operación que se imaginó que ocurría antes. A principios del siglo XX, los ingenieros reconocieron que la **retroalimentación negativa** también podría dar lugar a oscilaciones, pero es mucho más difícil determinar mediante un ensayo mental si un sistema de retroalimentación negativa generará oscilaciones sostenidas, ya que esto depende de las no linealidades en las interacciones. Hacerlo requiere un modelo matemático, que depende no solo de una descripción matemática adecuada del sistema, sino también de las herramientas adecuadas para realizar los cálculos necesarios. Inspirado por el mecanismo de retroalimentación descrito por Jacob y Monod (1961¹) y Goodwin (1965) se desarrolló una representación matemática², pero

utilizando una computadora analógica, se subestimó seriamente el valor necesario para que un parámetro crítico sostenga las oscilaciones. Cuando los investigadores circadianos identificaron un ciclo de retroalimentación negativa como un mecanismo candidato para los ritmos circadianos, revivió el modelo de Goodwin, pero incorporó retrasos para permitir oscilaciones sostenidas con un valor de parámetro más realista que el requerido por el modelo de Goodwin. La retroalimentación negativa e incluso positiva se invocó en el análisis de los sistemas biológicos en los últimos años del siglo XX, pero el descubrimiento de otros modos de organización quedó rezagado. Parecía que podría no haber ningún principio general que pudiera caracterizar los mecanismos biológicos complejos; cada uno podría tener que analizarse utilizando su propio modelo computacional. Si es así, la comprensión de la organización sería en el mejor de los casos un derivado de la investigación mecanicista dirigida a identificar partes y operaciones; después de que se identificaran, se podrían usar modelos matemáticos para determinar si el mecanismo podría generar el fenómeno en cuestión. Sin embargo, no se podía esperar anticipar la organización a partir del tipo de comportamiento exhibido por los fenómenos.

El avance más significativo aportado por la biología de sistemas son las nuevas herramientas para analizar los patrones de organización de los sistemas biológicos que permiten a los biólogos comenzar a aplicar ingeniería inversa a los sistemas biológicos³. Estas herramientas implican la aplicación y el desarrollo de recursos de la teoría de grafos, el modelado computacional y la teoría de sistemas dinámicos para comprender los sistemas biológicos. Al identificar redes del mundo nano y, además de notar su ocurrencia generalizada en sistemas naturales, señalaron su poder en el procesamiento de información.

Aunque se reconoce que muchas redes biológicas, incluidas las redes reguladoras de genes y las redes de interacción de proteínas, exhiben una organización en su mundo nano con una distribución de conexiones siguiendo por ejemplo la ley de potencias, ha demostrado ser útil para explicar propiedades como la robustez de los sistemas biológicos, tales análisis de la estructura de la red global todavía se encuentran en las primeras etapas de desarrollo. La esperanza es que a medida que avanza la investigación, se puedan identificar subcategorías dentro de la región del mundo nano y analizar sus propiedades dinámicas. Ese progreso se ha producido con respecto a las configuraciones locales dentro de las estructuras de red. Al identificar patrones

frecuentes de conexiones entre pequeñas cantidades de unidades, comúnmente conocidas como módulos, y luego desarrollar modelos para determinar los tipos de comportamientos a los que tales subgrafos darían lugar, han proporcionado recursos potentes para comprender cómo se comportarán los sistemas⁴ biológicos.

En las discusiones sobre biología de sistemas, a menudo se hace referencia a la **causalidad de arriba hacia abajo frente a la de abajo hacia arriba**. Para dar sentido a los fenómenos para los que se utilizan estos términos al tiempo que se restringe la causalidad a contextos intraniveles (acomodando las relaciones ascendentes y descendentes en términos de la relación constitucional entre partes y todos). Pero la teoría de grafos proporciona una forma potencialmente más informativa de visualizar lo que se ha pensado como procesos causales intra e interniveles mientras se descarta el sentido no bien articulado de los niveles. Lo que se necesita es un medio para detectar dentro de una red los casos en los que un conjunto de nodos constituyen un módulo que exhibe actividad dinámica endógena (por ejemplo, oscilación endógena sostenida) mientras aún está abierto a influencias de otros lugares. Dicho **módulo** es lo que podría haberse caracterizado como una entidad de nivel superior, pero en el grafo es simplemente un conjunto organizado de nodos. Si se desea, se puede dibujar un círculo alrededor del módulo para indicar que exhibe una dinámica endógena en la que cada componente está respondiendo a otros nodos en el módulo y la respuesta a cualquier entrada desde el exterior será modulada por la dinámica dentro del módulo. También es apropiado utilizar un círculo para los componentes del módulo, ya que, de hecho, las entidades correspondientes a la mayoría de los nodos dentro de una red pueden descomponerse ellas mismas en un conjunto de componentes organizados. El hecho de hacerlo dependerá de si los detalles de los procesos internos en las entidades correspondientes a los nodos se toman como importantes para la explicación buscada. Aunque un proceso causal afectará típicamente a algunos componentes de un módulo más que a otros, lo que luego afectará a otros componentes del módulo con el tiempo, puede ser suficiente para los propósitos explicativos de uno simplemente ver el proceso como que afecta al módulo. Dentro de esta perspectiva de red, la causalidad ascendente surge cuando los procesos causales locales se propagan dentro de un módulo para generar un sistema cuyas respuestas dependen de esos procesos locales. Sin embargo, estos mismos procesos causales son también los que median los efectos de arriba hacia abajo, ya que dan como resultado que los componentes de la red se comporten de manera diferente según el estado de la red en su conjunto. Esta

explicación se aplica iterativamente a medida que uno se mueve (a) hacia partes más locales de la red e identifica módulos con sus propios patrones sistémicos de organización en los que el comportamiento de las partes está determinado en gran medida por el comportamiento de otras partes o (b) hacia fuera a partes más globales en módulos que se forman al realizar tal organización entre módulos locales, desde la actividad dentro de los módulos locales es modulada por la actividad en otros módulos de la estructura mayor⁵.

Lo que es realmente fundamental en términos explicativos de la biología de sistemas no es un mero esfuerzo lógico, sino una empresa humana mediada por la tecnología (aunque no informada por ella). Este punto se vuelve particularmente interesante cuando se enfoca en explicaciones biológicas. Hay un proceso en curso, en las ciencias biomédicas en particular, que nos está obligando a revisar los supuestos epistemológicos a lo largo del progreso mismo de la comprensión científica.

¹ Auboeuf, Didier. (2021). The Physics–Biology continuum challenges darwinism: Evolution is directed by the homeostasis-dependent bidirectional relation between genome and phenotype. *Progress in Biophysics and Molecular Biology*. 10.1016/j.pbiomolbio.2021.05.008.

² Lakhova, Tatiana & Kazantsev, Fedor & Lashin, Sergey & Matushkin, Yury. (2021). The finding and researching algorithm for potentially oscillating enzymatic systems. *Vavilov Journal of Genetics and Breeding*. 25. 318-330. 10.18699/VJ21.035.

³ Green, Sara & Levy, Arnon & Bechtel, William. (2014). Design Sans Adaptation. *European Journal for Philosophy of Science*. 5. 15-29. 10.1007/s13194-014-0096-3.

⁴ Ali, Md & Parisutham, Vinuselvi & Choubey, Sandeep & Brewster, Robert. (2020). Inherent regulatory asymmetry emanating from network architecture in a prevalent autoregulatory motif. 10.7554/eLife.56517.sa1.

⁵ Green, Sara. (2017). Scale Dependency and Downward Causation in Biology. *Philosophy of Science*. 85. 10.1086/699758.

Módulo 0. El universitario y sus demonios

0. Los demonios que desafían la explicación racional científica

El vaso de la ciencia está medio vacío. Los investigadores de todo el mundo están obsesionados con todo lo que aún no sabemos. Era lo mismo hace cien años, y más de cien años antes también. De vez en cuando, el progreso llega. Algo hace clic en la cabeza de alguien. Todo cae en su lugar. El resultado es nada menos que la magia de la razón y la determinación humana. Lo que una vez había sido invisible de repente parece haberse estado escondiendo a la vista. La inspiración nos sucede a todos nosotros: escritores, poetas, científicos, profesores y estudiantes universitarios, así como a la gente en general. La brecha entre el conocimiento y la imaginación no es tan inescrutable como se ha hecho parecer. Es consecuencia mucho después del momento en que estas primeras ideas se evaporan y las preocupaciones prácticas toman el relevo.

Al previsualizar un mundo de maravillas mucho antes de que se abran las cortinas y comience el espectáculo, podemos sentarnos en el ensayo de nuestro propio futuro científico y tecnológico. La antesala del descubrimiento es un lugar donde las ideas se forjan antes de que vean la luz del día. Es la incubadora la que da forma a la ciencia antes de que se pruebe en tecnología. Cuando el espectáculo de nuestros logros incluye las evidencias, fundamentos y justificaciones que condujeron a ellas, provocan que el conocimiento se vea diferente. ¿Cómo podemos explicar la trayectoria de la ciencia y la tecnología que nos ha llevado de la máquina de vapor al microchip, de la revolución de la Inteligencia Artificial a la tecnología CRISPR (es una herramienta simple pero potente para editar genomas)?.

Un científico al amanecer en su laboratorio escribe manuscritos, entrena nuevos colegas, enseña a jóvenes universitarios, escribe libros... Este camino tiene una lógica clara que funciona de manera fragmentaria, sin embargo, en algún lugar del camino algo más grande que los propios actores, parece irrumpir. Los pensadores han quedado fascinados por los momentos de descubrimiento en la ciencia, cuando los científicos genios tienen una brillante idea: Eureka. Los avances a menudo llegan cuando menos se esperan. Lo que antes era imposible ya no lo es. Los nuevos experimentos y tecnologías se conciben por primera vez en la mente de los científicos. Son pensables mucho antes de que sean factibles. Cuando los científicos trabajan con

ideas radicales en caminos no explorados, sus mentes con frecuencia están en el espacio del laboratorio de las ideas, hechas de piezas de matemáticas y conexiones aportadas por toda la comunidad científica.

La naturaleza sorprendente del conocimiento y la invención puede llevarnos a sospechar que algo parecido a una fuerza inconsciente convive detrás de los límites de la razón e impulsa su desarrollo desde el exterior. Descubrir tiene su propia historia retorcida, fascinante y a veces aterradora. También tiene sentido su propio vocabulario técnico altamente desarrollado. Los científicos a menudo usan la palabra “demonio” dentro de las frases más preliminares de su investigación. Designa algo que aún no se conoce o se entiende completamente. Estos demonios no son religiosos, sobrenaturales, monstruosos o simplemente duendes malvados. Se refieren a algo que desafía la explicación racional y puede desconcertar o romper una hipótesis o una ley de la naturaleza. Su función no es principalmente metafórica o figurativa. Son términos técnicos cuya definición se puede encontrar en casi cualquier diccionario. El Oxford English Dictionary define “dominio” en la ciencia como: “cualquiera de varias entidades de noción que tienen habilidades especiales, utilizadas en experimentos de pensamiento científico”. Se menciona con referencia a la persona particular asociada con el diseño de experimentos que sigue un patrón originado con René Descartes, el pensador del siglo XVII conocido por inaugurar la era de la razón. El dominio de Descartes abrió las puertas a muchos otros, continuando hasta el día de hoy. Los nuevos nombres se añaden tan pronto como pasan a formar parte del argot del laboratorio. El demonio de Laplace, siguió los pasos del demonio de Descartes y se convirtió en un modelo para nuevas máquinas calculadoras que potencialmente podrían determinar la posición precisa y el movimiento de todas las partículas en el universo para conocer todo el pasado e incluso el futuro.

Estos dos demonios pronto se enfrentaron a la dura competencia de la criatura victoriana llamada “el demonio de Maxwell”, que podría causar estragos con el curso habitual de la naturaleza. A medida que la ciencia crecía en prestigio y complejidad, muchos otros demonios fueron invocados y nombrados en honor a Charles Darwin, Albert Einstein, Max Planck, Richard Feynman y otros.

Averiguar las cosas a menudo implica invocar a los demonios como una categoría útil para articular y llenar los vacíos del conocimiento existente. Cuando se enfrentan a un

problema particularmente difícil, o cuando el universo no está funcionando de la manera que uno esperaría, los científicos inmediatamente comienzan a sospechar de un perpetrador. Además de recibir el apellido de científico que comienza a pensar en el enigma, los culpables a menudo son antropomorfizados a medida que se convierte en planos para las tecnologías futuras. Los investigadores a veces se refieren a ellos como él, otras veces como ella, y a menudo como ella. A medida que los científicos imaginan demonios con habilidades que compiten y los imaginan colaborando o luchando entre sí, inspiran a la creación de los arreglos tecnológicos cada vez más complejos. Los prototipos son constantemente actualizados. Las nuevas versiones están a la vuelta de la esquina, pronto serán lanzados.

Al realizar una búsqueda de información entre archivos de una computadora, se utilizan líneas de código que son denominadas “demonios” para encontrar las coincidencias que está buscando. Cuando inicia sesión en Internet o usa un teléfono inteligente, miles de demonios de este tipo (programas) se ponen a trabajar para suavizar los procesos de comunicación entre usted, su dispositivo y los dispositivos de los demás. Hoy en día, estos demonios son fundamentales para infraestructura de comunicaciones que nos rodean. Los mecanismos internos de interacción, causalidad, funcionalidad, correlación..., en su conjunto son esos demonios que se buscan revelar, que se nos esconden y resisten a nuestras observaciones y diseños experimentales.

El poeta francés del siglo XIX Charles Baudelaire le dijo a los científicos que dibujaron ecuaciones en las ventanas por las que miraron el mundo, “el mejor truco del demonio es persuadirnos de que no existe”. De las innumerables cosas que las matemáticas nos permiten modelar en la biología, la química o la física, ¿qué hemos hecho? Un paseo por la historia de la ciencia y la tecnología nos muestra que las innovaciones a menudo engendran arrepentimiento, la determinación puede convertirse en retorcimiento de manos y el euforia inicial de la búsqueda de la conciencia humana. La literatura de la historia de la ciencia está llena de memorias de retrospectivas escritas por científicos que se enfrentaron a la misma pregunta después de ver cómo se había puesto en uso sus investigaciones. El conocimiento da poder, dejándonos hacer frente a la complicada elección ética de que hacer entre bien y el mal. Incluso nuestras tecnologías más avanzadas no nos han traído todas las promesas que esperábamos. Vivimos con temor de que las innovaciones más preciadas en ciencia y tecnología puedan caer en las manos de mentes de terror. Hemos sido testigos de desarrollos virtuosos y honorables,

que rápidamente fueron usados con fines destructivos, para causar el mayor dolor posible.

Todo lo que se necesita para convertir algo bueno en algo terrible es una dosis ligeramente de populismo político de izquierda o derecha, un aumento incremental en la cantidad o un cambio imperceptible de contexto. Los pesticidas se han utilizado en cámaras de gas contra personas inocentes, los fertilizantes se pueden usar para construir bombas, los cohetes espaciales pueden entregar armas de destrucción masiva, las vacunas se adaptan fácilmente para la guerra biológica, la cura para enfermedades genéticas pueden convertirse en la base de intervenciones eugenésicas, el mismo instrumento se puede usar para curar o lastimar, y así sucesivamente. Lo que una vez fue una solución puede convertirse en una herramienta para perpetrar un crimen. Un sueño puede convertirse en una pesadilla en una abrir y cerrar de ojos.

El panorama del desarrollo tecnológico que emerge no es del todo bueno. La espada del conocimiento corta dos caminos. Hemos pensado en los peligros del conocimiento de esta manera desde que apareció por primera vez como un concepto en la historia. El relato bíblico de la expulsión del edén, describe al conocimiento como algo trasgresor e incluso demoníaco. Una criatura asociada con el Diablo, más astuta que cualquiera de los otros animales salvajes, tienta a Adán y Eva a morder la fruta prohibida (el conocimiento).

Hasta el día de hoy, un deseo de adquirir conocimiento, de obtener sabiduría, continúa siendo considerado trasgresor y, a veces, incluso pecaminoso. ¿Por qué creemos que la curiosidad mató al gato? En otras palabras, ¿hay algo en la búsqueda del conocimiento que casi siempre es matar una idea en nosotros que no está justificada? La historia de los demonios en la ciencia, nos permiten ver algo que la mayoría de las historias sociales pasan por alto, el arco de la ciencia y la tecnología moderna, tiene que ver con las limitaciones de nuestros sentidos y el poder de la razón. Estos estudios fueron fundamentales para el desarrollo de la mecánica, la termodinámica, la relatividad, la cuántica y la cosmología. Muchas tecnologías se consideran mágicas y fantásticas cuando están en papel, son consideradas demoníacas; esto solo es por un corto periodo en lo que la sociedad asimila el periodo del tiempo que le toco vivir, algo completamente nuevo. Es solo cuando la investigación es nueva, innovadora, misteriosa y potencialmente transformadora de la imaginación colectiva.

Los castillos en el cielo rara vez están vacíos. Hemos domesticado al electrón, y se le llama electrónica. Nuestra imaginación hace maravillas, y muchos estudiosos se han dedicado a estudiarla. Sin embargo, a menudo se asume que su papel en la ciencia es secundario. Tradicionalmente se considera que es un arte privado, demasiado rebelde para estudiar, fuera de los límites de la investigación racional, incipiente, resbaladizo, oscuro y tal vez incluso la erudición sobre los experimentos mentales ha crecido en los últimos años, la mayoría de los estudiosos todavía realizados en laboratorios y centro de investigación. El papel de la imaginación en la ciencia sigue siendo retratado como un inconveniente fuera del laboratorio y se cuela astutamente y ocasionalmente, como un vergonzoso hermano o hijo de las artes y las humanidades que aparece sin invitación. Pero su poder no se detiene cuando los científicos entran en el laboratorio o escriben sus ecuaciones. Toda la empresa de la ciencia, desde la teoría hasta el experimento y la comunicación pública, está completamente impregnada de nuestra imaginación. Cuando pensamos, razonamos y tomamos decisiones, simultáneamente pensamos en el futuro, mucho y más allá.

El rompecabezas se rompe de la siguiente manera, cada ser vivo que conocemos surgió de otro ser vivo, sin embargo, tenemos razones para pensar que no había vida en absoluto cuando el universo se puso en marcha por primera vez. Esta observación implica que la primera vida allí creció a partir de cosas que no estaban vivas, y la pregunta es si este gran evento ocurrió misteriosamente, de una manera incomprensible para las leyes de la física y la química tal y como las conocemos, o que tuvo lugar en una serie de pasos que podemos entender. En otras palabras, ¿cómo comenzó la vida? Los científicos, los observadores curiosos del mundo natural y muchos otros tipos de buscadores contemplativos que luchan con la condición humana pueden estar de acuerdo en la importancia de esta pregunta, pero reina mucha confusión y desacuerdo sobre qué tipo de respuesta estamos buscando.

A algunos biofísicos les gustaría saber qué tipos específicos de átomos chocaron entre sí para formar las primeras biomoléculas, mientras que otros dan por sentado que algo como esto sucedió, y centrar su investigación en su lugar en juzgar la probabilidad de tal evento bajo diversas condiciones. Para el caso, ¿qué cuenta como vida y qué no? ¿Fue la presencia temprana de la vida en el mundo un continuo gris a través del cual las cosas progresaron gradualmente con el tiempo, o insistimos en que debe haber

habido un momento en que irrumpió decisivamente en la escena? Si es lo primero, ¿con qué precisión podemos diferenciar entre algo que está vivo y algo que no lo está? Si este último, teniendo en cuenta que la vida es buena para hacer un número impresionante de cosas, ¿cuál de ellas nació primero? ¿Estaba garantizando hacerlo, de la misma manera que el hielo siempre se derrite bajo el cálido sol, o debe haber sido provocado por un suceso extraño, uno tan raro que nuestra mejor teoría de cómo sucedieron las cosas no es ninguna explicación en absoluto? Entendemos algo de por qué el hielo se forma cuando el agua se enfría lo suficiente, o por qué las estrellas se encienden cuando la gravedad es lo suficientemente fuerte como para exprimir el gas hidrógeno, pero es extraña y maravillosamente más difícil articular las condiciones físicas en las que se garantiza que la materia no está viva. Al igual que el espectador escéptico en un espectáculo de magia, muchos de nosotros simplemente no estaremos satisfechos hasta que llegemos a ver cómo se hizo el truco.

La investigación científica tenaz no es, sin embargo, la única razón por la que la gente anhela tanto saber algo de cómo comenzó la vida. Un ser humano reflexivo, de hecho, incluso un científico dispuesto a mirarse honestamente en el espejo y examinar sus propios sentimientos cuidadosamente, puede admitir que la razón por la que la pregunta tiene tal dominio sobre nosotros es que expresa un anhelo compartido. La búsqueda de significado y propósito comienza con preguntarse de dónde venimos y de qué somos parte, y buscar un relato de cómo comenzaron las cosas, en parte porque la gente usa ideas sobre el pasado para decidir cómo actuar en el presente y en el futuro.

Si de todo lo que los seres humanos están hechos se puede encontrar en un estanque o un conjunto de química, ¿entonces qué es lo que realmente nos hace diferentes? ¿Son los humanos simplemente animales, algo más? ¿Expresa nuestra existencia la intensión de un creador que nos diseñó a su imagen, o como nosotros y todo la vida meramente una variedad exótica de escarcha condensada en la capa delgada entre el suelo y el cielo?

Los seres vivos logran una variedad de hazañas que, aunque no son exclusivas de la vida, son ciertamente distintivas de ella. Por ejemplo, hacen copias de sí mismos, consumen combustible, y predicen con precisión el medio ambiente circundante. Todos estos procesos son parte integral de lo que es estar vivo, y cada uno de ellos puede ser estudiado sistemáticamente desde la perspectiva de la termodinámica.

Haciendo hincapié en el progreso reciente en una rama de rápido crecimiento de la termodinámica. Haciendo hincapié en el progreso de la termodinámica conocida como mecánica estadística lejos del equilibrio. El punto clave es darnos cuenta de que, así como los seres vivos tienen propiedades especializadas determinadas por sus genes que han heredado de sus antepasados, también las colecciones de partículas que interactúan físicamente tienen propiedades especializadas que provienen de las formas pasadas en la que se han ensamblado. Al ser continuamente empujadas y golpeadas por los patrones presentados en el medio ambiente, la materia puede someterse a una exploración continua del espacio de posibles formas cuyo ritmo y forma se ajustan a esos patrones de maneras que se parecen mucho a la vida. Incluso más que la mayoría de los otros temas científicos, este seguramente exige una conversación más amplia. Y sea porque uno necesita lidiar poderosamente con la cuestión simple de qué es estar vivo y lo que no está, o por qué los argumentos sobre lo excepcional, el valor y el propósito de la vida, constituyen una gran parte de lo que la gente ha estado en desacuerdo y por lo que lucha a lo largo de los siglos, parece completamente necesario poner nuestro examen de la frontera entre la vida y la no vida en un contexto científico convenientemente rico. Puede que haya más de una forma de hacerlo, en principio, pero aquí es donde nuestros propios compromisos personales profundamente intervienen en el esfuerzo de pensar.

0.1 La biología en las nanomáquinas celulares

A menudo el pensamiento reduccionista falla¹, en estos casos, las leyes de la física en la biología que supuestamente la gobiernan, todo está inundado con la inmensidad de lo que no sabemos, no podemos mentir, o carecemos de la capacidad de computar directamente una inmensa cantidad de información. Se trabaja en tales sistemas, pero desde la física con ecuaciones fundamentales que gobiernan las partes microscópicas y desde la biología, con situaciones de información incompleta empleado filosofía de Bayes.

Hay algo obviamente razonable en la siguiente noción: si toda la vida se construye a partir de átomos que obedecen a ecuaciones precisas que conocemos, lo que parece ser cierto, entonces la existencia de la vida podría ser una consecuencia posterior de estas leyes que aún no hemos llegado a calcular. Esta es esencialmente la forma de pensar de un físico, y para su crédito, ya ha hecho mucho para ayudarnos a comprender cómo

funcionan los seres vivos. Gracias a pioneros como Max Delbrück, que pasó de la física a la biología a mediados del siglo XX, la influencia de los análisis cuantitativos de las ciencias físicas ayudó a dar lugar a enfoques mecanicistas y moleculares en la biología celular y bioquímica que condujeron a muchos descubrimientos revolucionarios². Las técnicas de imagen como la cristalografía de rayos X, la resonancia magnética nuclear y la microscopía de superresolución han proporcionado un retrato vivido del ADN, las proteínas y otras estructuras más pequeñas que una sola célula que hacen que la vida funcione a escala molecular³.

Además, la descifrar el código genético, nos hemos vuelto capaces de aprovechar la maquinaria de las células vivas para cumplir nuestros comandos mediante el ensamblaje de nuevas macromoléculas de nuestra propia creación. A medida que hemos obtenido una imagen cada vez más precisa de cómo los bloques de construcción más pequeños y simples de la vida encajan para formar el todo, se ha vuelto cada vez más tentador imaginar que los rompecabezas más difíciles de la biología solo pueden resolverse una vez que descubramos cómo abordarlos en términos de física de mecanismos causales.

Pero para abordar el tema de la vida con esta actitud nos podría fallar, por al menos dos razones. La **primera razón** podríamos llamarla la falacia del reduccionismo. El reduccionismo es la presunción de que cualquier pedazo del universo que podríamos elegir estudiar funciona como un espécimen de relojería antigua, de modo que es fácil o al menos, eminentemente posible predecir el comportamiento del todo una vez que sus partes empujan y se mueven con las demás. El sueño de explicar y predecir todo a partir de unas pocas reglas simples ha capturado durante mucho tiempo la imaginación de muchos científicos, particularmente los físicos. Y, para ser justos, una gran cantidad de buena ciencia ha sido impulsada por el hambre de algunos investigadores de una explicación más completamente reductiva del fenómeno que les interesa. Después de todo, hay cosas en el mundo que pueden entenderse como el resultado de interacciones conocidas entre varias piezas más simples. Desde el aumento y la caída de las mareas oceánicas con el tirón gravitacional de la luna, hasta la forma en que algunas enfermedades genéticas se pueden rastrear hasta eventos moleculares que surgen de la química alterada de un pequeño parche en la superficie de una proteína, a veces lo que estamos estudiando parece una suma comprensible de sus partes.

El segundo error en la forma que las personas han visto el límite entre la vida y la no vida, se origina en la forma en que usamos el lenguaje. Muchas personas imaginan que si entendemos la física lo suficientemente bien, eventualmente comprenderemos lo que es la vida como un fenómeno físico de la misma manera que ahora entendemos cómo y por qué el agua se congela o hierve. De hecho, a menudo parece que la gente espera que una teoría física lo suficientemente buena podría convertirse en el nuevo estándar de oro para decir lo que está vivo y lo que no. Lo que argumentamos aquí, sin embargo, es que este enfoque no reconoce que nuestro propio papel en dar nombres a los fenómenos del mundo precede a nuestra capacidad de decir con claridad lo que significa incluso llamar vivo a algo. Un físico que quiere idear teorías de cómo se comportan o emergen los seres vivos de la materia más elemental, tienen que comenzar por tomar decisiones intuitivas sobre cómo traducir las características de los ejemplos de vida que conocemos a un lenguaje físico. Después de que uno lo ha hecho, rápidamente queda claro que el límite entre lo que está vivo y lo que no lo está es algo que ya se dibujó al principio, a través de una forma diferente de hablar que la que proporciona la física. Por lo tanto, el objetivo adecuado para el relato de las cosas de un físico debería ser encontrar una manera de describir ese límite en términos físicos precisos, de modo que podamos obtener una nueva visión de cómo se podría llegar a mover la materia de un lado a otro del límite.

Hasta cierto punto, una inclinación esperanzadora hacia el reduccionismo se expresa en la pregunta misma de dónde viene la vida. Miramos a un organismo vivo y no podemos evitar preguntarnos si tal éxito impresionante en forma y función podría ser simplemente el resultado de un montón de piezas más básicas que rebotan entre sí como bolas de billar simples y predecibles. ¿Hay algo más en la máquina aparte de todas sus partes torpemente vibrantes? Si no lo hay, ¿no debería eso significar que eventualmente podemos entender cómo encaja todo el asunto? Dicho de otra manera, ¿no tendría que dividirse todo en una serie de pasos racionalizados, donde cada uno de los siguientes sigue de manera sensata y predecible de la última? Si es así, ¿cómo es que no es lo mismo decir que queremos reducir la vida a una actuación coreografiada dirigida por un conjunto simple y calculable de reglas físicas conocidas?

Hay que reconocer que los físicos ya han identificado algunas reglas que demuestran hacer predicciones altamente precisas en sistemas que alguna vez parecieron

irremediable y misteriosamente complicados. Gracias a las ideas de personas como Kepler, Newton, Galileo, el movimiento de los cuerpos celestes es ahora un libro abierto, y nuestra capacidad para calcular a dónde van estas luces brillantes en el cielo es una banalidad tan poco notable que ahora es posible obtenerla en una amplia educación en física en muchas grandes universidades sin profundizar en la especialización de la mecánica orbital rigurosa. Imagínese, sin embargo, a un brillante filósofo natural en cualquier momento durante la mayor parte de la historia humana, y maravillarse de la complejidad aparentemente intratable de cómo el sol, la luna y las estrellas parecen reorganizarse continuamente en el firmamento como lo hacen todos días. La idea de que un par de ecuaciones concisas que describen la gravitación y el movimiento bajo la fuerza podría atraer galaxias distantes, los planetas errantes y las cajas colgadas de resortes en espiral en un marco teórico integral debería ser inconcebible incluso para el genio más grande de cada era durante miles de años. El alcance y la importancia de la revolución que comenzó con Newton y sus contemporáneos son difíciles de exagerar su enorme calado en la civilización.

0.2 A pasos de mayor complejidad

¡Llego el siglo XX de Einstein!, comenzó contemplando las ecuaciones que describen el movimiento de La Luz, y a través de la pura fuerza de perspicacia terminó reimaginando los orígenes de la gravedad, para finalmente explicar el último rompecabezas restante del movimiento planetario que Newton no podía tocar (a saber, Mercurio). Mientras tanto, la ecuación de onda mecánica cuántica de Erwin Schrödinger desbloqueó el átomo, proporcionando una elegante explicación cuantitativa para los colores de luz emitida por varios tipos de gas electrificados. Esta era una teoría extraña y poco intuitiva del funcionamiento interno matemático de objetos demasiado pequeños para ser vistos y tocados en su individualidad, pero aún podía igualar las mediciones experimentales con una precisión asombrosa. A raíz de estas grandes victorias científicas, uno podría perdonar a uno o dos científicos extraños por sentir que toda la imprevisibilidad podría eventualmente ser barrida a medida que llegaban teorías más nuevas y cada vez más brillantes.

Sin embargo, en una inspección más cercana, este desfile de victorias para la ciencia teórica reductiva revela cierto sesgo. La que estos y muchos otros ejemplos de teorías físicas exitosas tienen en común es que funcionan mejor cuando se trata de predecir un

pedazo bien asilado del mundo descrito por una formulación matemática relativamente simple que involucra algunas cosas diferentes que uno puede medir: el sistema solar de un planeta, el átomo de hidrógeno único y solitario, y así sucesivamente. En cada uno de estos casos, la teoría sucede por filtrado del resto del universo y centrándose en una pocas ecuaciones que describen con precisión las relaciones entre un pequeño número de cantidades físicas. De la misma manera, mirar un poco más de cerca estos mismos modelos nos da una idea de dónde vas a flaquear el pensamiento reductivo.

A veces nuestros modelos nos decepcionan porque lo logran captar algo fundamental para el comportamiento del sistema que están tratando de describir. Por ejemplo, como se mencionó anteriormente, la gravedad del sol es sentida tan fuertemente por el planeta Mercurio que la Ley de Gravitación de Newton fue notoriamente insuficiente para predecir la forma observada de la órbita; solo con la sutiliza añadida de la Relatividad General se podría tapar ese agujero. Aún así, hay muchas situaciones en campos gravitacionales más débiles donde Newton trabaja de manera bastante adecuada, en principio. Sorprendentemente, sin embargo, incluso cuando el modelo que estamos utilizando es una declaración perfectamente precisa de las reglas por las cuales se comporta el sistema de interés, el problema de la predicción puede ser resbaladizo.

Stephen Edelston Toulmin argumentó defender una afirmación hecha al afirmar algo⁴. Observó ciertas características invariantes de campo de nuestro saber. Toulmin no se preocupa por cómo lo hacemos o deberíamos llegar a él. Comenzamos reconociendo a varios candidatos de conocimiento para una solución, candidatos que son "posibles" en el sentido de que tienen derecho a ser considerados. Luego consideramos la información que tenemos a nuestra disposición sobre estas sugerencias, quizás concluyendo que algunas son después de todo "imposibles", quizás identificando una como la más "probable" en el sentido de ser más merecedora de aceptación, quizás identificando una como "presumiblemente" correcto a menos que se apliquen ciertas condiciones inusuales o excepcionales. Durante este proceso de justificación racional, lanzamos lo que Toulmin llamó "microargumentos", para los cuales propuso un patrón de análisis invariante en el campo diseñado para hacer justicia al proceso de defensa de un reclamar contra un interlocutor. Este patrón, que ha llegado a conocerse como

"modelo de Toulmin" o "esquema de Toulmin", difería radicalmente del análisis lógico tradicional de un microargumento en premisas y conclusión.

Ante el desafío de defender nuestro reclamo por un interrogador que pregunta: "¿Qué tienes para continuar?", Apelamos a los hechos relevantes a nuestra disposición, que Toulmin llama nuestros datos (D). Puede resultar necesario establecer la exactitud de estos hechos en un argumento preliminar. Pero su aceptación por parte del interlocutor, ya sea inmediata o indirecta, no necesariamente pone fin a la defensa. Porque el retador puede preguntar sobre la relación de nuestros datos con nuestra afirmación: "¿Cómo se llega allí?" Nuestra respuesta, en su forma más clara, tomará la forma: "Datos como D dan derecho a sacar conclusiones o hacer afirmaciones, como C". Una proposición de esta forma Toulmin llama una orden legal (W). Las garantías, señala, confieren diferentes grados de fuerza a las conclusiones que justifican, lo que puede indicarse calificando nuestra conclusión con un calificativo (Q) como "necesariamente", "probablemente" o "presumiblemente". En el último caso, es posible que debamos mencionar las condiciones de la refutación (R) "que indiquen circunstancias en las que la autoridad de la orden tendría que dejarse de lado". Nuestra tarea, sin embargo, todavía no está necesariamente terminada. Porque nuestro interlocutor puede cuestionar la aceptabilidad general de nuestra orden legal: "¿Por qué piensas eso?" Toulmin llama a nuestra respuesta a esta pregunta nuestro respaldo (B). Destaca las grandes diferencias de tipo entre los soportes en diferentes campos. Las garantías pueden defenderse apelando a un sistema de clasificación taxonómica, a un estatuto, a las estadísticas. Es esta diferencia en el respaldo lo que constituye la dependencia de campo de nuestros estándares de argumentación. En última instancia, todos los microargumentos dependen de la combinación de datos y respaldo. En casos excepcionales, verificar el respaldo implicará verificar el reclamo; Toulmin llama a estos argumentos "argumentos analíticos". La mayoría de los argumentos no son de este tipo, por lo que los criterios puramente formales no son suficientes para su evaluación; Toulmin los llama "argumentos sustanciales". El tipo de respaldo que es aceptable para un argumento sustancial dado dependerá del campo al que pertenece. Para ilustrar la contribución de estos componentes Toulmin enfatizó una serie de puntos que ahora son familiares, pero que aún merecen atención⁵:

1. El razonamiento y la argumentación implican no solo el apoyo a los puntos de vista, sino también el ataque contra ellos.

2. El razonamiento puede tener conclusiones matizadas.
3. Hay otros buenos tipos de argumentos además de los de la lógica formal estándar.
4. Los supuestos no expresados que vinculan premisas con una conclusión se consideran mejor como licencias de inferencia que como premisas implícitas.
5. Los estándares de razonamiento pueden depender del campo y pueden ser ellos mismos objeto de argumentación.

0.3 La predictibilidad y lo ontológico como generadores de tecnología

El pensamiento científico moderno: pensar desde la teoría y librarnos de los demonios con audaces diseños experimentales. Este modelo de pensamiento nos pone en el camino de darnos la oportunidad de escribir, pensar y debatir sobre el tema de colaboración entre teóricos y empiristas. Es esencial para hacer sólido nuestro pensamiento. En la universidad estamos lidiando con estos difíciles problemas intelectuales, intentamos animar un diálogo sólido entre diseñadores experimentales, teóricos e ingenieros. Nuestra experiencia nos dice que la comunidad intelectual solo puede dar pasos éticos progresistas si logra emprender discusiones y debates honestos y flexibles entre teóricos y empiristas. Este camino intenta aclarar nuestro pensamiento sobre la forma en que hacemos investigación.

En primer lugar escribimos para estudiantes universitarios que están pensando en cómo encajan la investigación científica, la ingeniería y el diseño experimental. Es fácil perderse en el bosque solo mirando los árboles durante el estudio de asignaturas del plan de estudios. En medio de la formación científica formal en clases de metodología o construcción de teorías, es difícil ver cómo estas empresas se conectan con la ciencia de datos y la producción de tecnología. En segundo lugar, la comunidad académica de profesores e investigadores, puede ayudar a comprender cómo sería benéfico resolver el divorcio entre empíricos y teóricos. Hacer preguntas como modelos (los teóricos) y los empíricos trabajan juntos para avanzar en objetos disciplinarios y la creación de nuevas tecnologías. Pero además, discutir este tema, es un apasionante placer para tomar café y escribir con un interés en la **teoría formal y la inferencia causal**, pensar en el valor de los modelos abstractos como generadores de tecnología y nuevos diseños biotecnológicos.

Discutir los papeles del modelador y el empírico será útil para que los estudiantes universitarios en la búsqueda de la excelencia, lean un marco moderno en el que se desarrolla la ciencia de datos. Creemos que hacerlo clarifica la interacción entre la teoría y lo empírico. Nuestro enfoque es la discusión de las ideas manteniendo al mínimo los tecnicismos, buscando la comprensión conceptual.

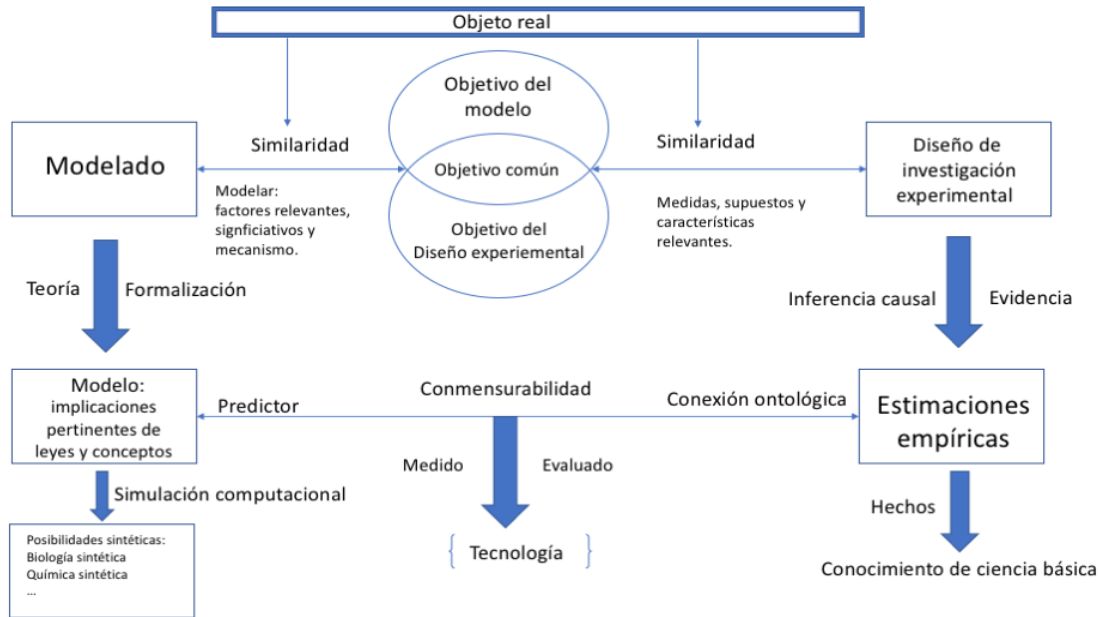


Figura 0.1 Conexiones entre modelado, diseño experimental y tecnología (apoyado en el trabajo de Quentin Ruyant sobre “Theories, Models and Representation⁶”)

El auge de la teoría formal y la revolución de la credibilidad son dos de los grandes desarrollos de la ciencia de los últimos 50 años. Con estos avances, el potencial para el diálogo productivo entre teóricos y empíricos nunca ha sido mayor. Por lo tanto, es penoso que, en las ciencias y en la ingeniería parezca mantener este hecho a la deriva. Irónicamente, estos dos desarrollos, deberían unirse más en la formación de científicos en las universidades, estar uniéndolos a los ingenieros, en cambio los está dividiendo. La revolución de la credibilidad ha sensibilizado desde la ingeniería a los científicos sobre los desafíos de separar: la correlación de la causalidad; y nos ha obligado a contar con la verosimilitud de las interpretaciones causales que se dan rutinariamente a las estimaciones empíricas⁷.

Por un lado, a algunos científicos les preocupa que la búsqueda de estimaciones causales creíbles esté desplazando el objetivo canónico de comprender fenómenos científicos importantes. ¿Por qué sucede esto? Muchos ven a los partidarios de la revolución científica de los datos en la credibilidad, desdeñosos de lo que puede ser y de hecho, se puede aprender por estudios empíricos de otros enfoques. La revolución de la credibilidad nos distrae de nuestros verdaderos objetivos. Esarey y Summer describen el auge del “empirismo radical” como divorciado de la teoría⁸. Otros se lamentan de que debido a no creer que una buena identificación causal, no siempre es posible en cuestiones de importancia central, la revolución de la credibilidad ha llevado a los científicos a reducir excesivamente la gama de preguntas que hacen. Al priorizar las estrategias de identificación de hechos, se corre el riesgo de perder de vista los intereses teóricos y analíticos que motivan la investigación científica.

En el otro lado están los científicos que han abrazado la revolución de la credibilidad, argumentando que gran parte del trabajo cuantitativo canónico en ciencias ofrece solo la ilusión de aprender sin un modelado. Para estos, no tiene sentido abordar preguntas que no se pueden responder con claridad. En su lugar, sugieren centrarnos en cuestiones accesibles a diseños de investigaciones creíbles. Justo antes de la revolución de la credibilidad como producción en masa de pseudo-hechos cuantitativos a través de regresión múltiple y otros medios de la ciencia de datos; Hauke argumenta que sin un diseño de un experimento cercano a lo natural, una discontinuidad o algún otro diseño, ninguna cantidad de modelado estadístico puede hacer que el paso de la correlación a la causalidad sea persuasivo⁹. Esta conclusión tiene implicaciones para el tipo de preguntas causales que somos capaces de responder con cierto rigor.

Esta cisma recuerda la división anterior abierta por el auge de la teoría formal y su mayor enfoque en la construcción de modelos. Al igual que los partidarios de la revolución de la credibilidad, los primeros teóricos de la elección racional, como Green y Shapiro describen: no sostenemos que los científicos tradicionales han estudiado los fenómenos equivocados, sino que han estudiando los fenómenos correctos de la manera equivocada¹⁰.

En el papel de los críticos actuales de la revolución de la credibilidad, están aquellos preocupados de una fetichización de la elegancia matemática, porque argumentan que distrae a los científicos del objeto de generar ideas genuinamente útiles para explicar o

adecuar la evaluación empírica. Muchos defienden que el progreso empírico se ha visto retrasado por lo que se denomina **investigación basada en métodos**, en lugar de realizar investigaciones basadas en problemas. ¿Qué es interesante o útil?, preguntaron los críticos, acerca de los modelos estrechos construidos sobre suposiciones que llevan, en el mejor de los casos, solo a un resultado en una relación distante con la realidad. Por ejemplo, en su crítica a los modelos formales, Fernández (2021) argumentó que son una idea consistente, precisa, pero trivial al menos en el valor de una nueva conjetura audaz que nos ayude a comprender algún problema importante del mundo real... una teoría lógicamente consistente pero empíricamente falsa que tiene poco valor¹¹.

La formalización (modelado matemático) permite a los científicos evitar errores de lógica y lograr una mayor transparencia. Pero estrecha las conexiones entre las suposiciones y las conclusiones fuera de todo lo que la teoría formal tuviera que ofrecer, esta sería una contribución muy importante. Algunos investigadores señalan tres virtudes de la formalización¹²: ver con claridad qué supuestos impulsan qué resultados; evitar errores de lógica a través del rigor y, lograr una especie de unidad o coherencia evitando hipótesis que dependen de supuestos contradictorios.

Esta cisma presagiaba la división cada vez más profunda de hoy entre teóricos y empíricos. Si bien, en principio, casi todo el mundo está de acuerdo en que los teóricos y empíricos deben trabajar juntos, en la práctica, cada parte siente que la otra a menudo no sostiene su parte del trato. Por un lado, un grupo de pensadores de mentalidad teórica que está desconcertado y consternado por el giro empírico hacia los diseños de investigación para responder de manera creíble a preguntas causales estrechas¹³. ¿Qué se preguntan los empiristas?, están obsesionados con responder cuidadosamente preguntas poco interesantes, en lugar de hacer un trabajo que hable de preguntas teóricas. Por otro lado, un grupo de pensadores con mentalidad empírica están igualmente desconcertados y consternados por el enfoque de los teóricos en modelos abstractos, desde su perspectiva son suposiciones demostrablemente falsas¹⁴. De qué sirven, se preguntan, ¿pueden ser útiles tales modelos para explicar el mundo o guiar la investigación empírica?

Como resultado de esta mutua perplejidad e insatisfacción, los grupos se están separando, yendo sobre sus propios asuntos y viendo al otro con creciente

escepticismo. Esta brecha cada vez mayor amenaza el vínculo entre teoría y lo empírico que son juntos esencialmente la empresa científica.

Ha llegado el momento de volver a unir estos dos grupos. La teoría formal y la revolución de la creatividad (ciencia de datos), son socios naturales que, juntas, pueden apoyar un diálogo en la formación de jóvenes universitarios. Una formación STEM entre teoría y lo empírico que nunca antes ha sido posible en la ciencia y el humanismo en un México de tradición negacionista, denunciado en 1928 por el nicolaita Samuel Ramos.

0.4 Darse la mano empiristas y teóricos

Sin embargo, impulsar el reconciliamiento de las disciplinas del oficio teórico y empírico, actualmente requiere integrar su potencial como entidades unidas. Empiristas y teóricos por igual impactan en la empresa del otro. Todos necesitan un marco mejor para encajar en la actividad universitaria de investigación y formación académica de nuevos científicos. Cada parte necesita comprender mejor qué tipo de conocimiento está tratando de crear la otra parte y cómo lo hacen. Solo con esta comprensión los teóricos verán cómo hacer que sus modelos sean genuinamente útiles para la investigación empírica. Y los empiristas verán cómo estructurar su investigación de manera que hablen de preguntas teóricamente significativas.

Si vamos a aprender a detectar oportunidades para una conexión más profunda entre teóricos y empíricos, tenemos que dedicar algún tiempo a los detalles. Detalles sobre por qué los teóricos escriben los tipos de modelos que hacen, y cómo esos modelos se pueden hacer más útiles para los empiristas. Detalles sobre por qué los empiristas usan los tipos de diseños de investigación que emplean, y cómo esos diseños de investigación se pueden hacer para dirigirnos mejor a preguntas teóricamente significativas. Detalles sobre cómo funciona exactamente la conexión entre ellos y qué hace a esa conexión informativa.

Pero sería un error empezar por ahí. Nos preocupa que si nos metemos de nuevo en los detalles, usted no tendría ningún contexto para entender por qué hacemos hincapié en alguno temas y restar importancia a otros. O cómo nuestra discusión de los detalles encaja en el panorama general, creado el diálogo sólido entre la teoría y los

empíricos que es necesario para una ciencia próspera. En síntesis, extrañaremos el bosque para ir a ver los árboles.

Es necesario un marco conceptual para pensar en cómo encaja toda esta empresa científica. En términos generales, ¿cómo funcionan los modelos teóricos y la investigación empírica?, ¿cómo los diseños experimentales o modelos se relacionan con el mundo real? Cuando comparamos una implicación teórica con un hallazgo empírico, ¿de qué tipo de afirmaciones científicas estamos aprendiendo? ¿Cómo podemos hacer tales comparaciones máximamente informativas sobre las preguntas que nos motivan, es decir, preguntas sobre cómo funciona el mundo?

Comparar y contrastar los enfoques prominentes para vincular la estructura y el análisis de mediación causal. Esa comparación nos lleva a una discusión de la perspectiva de nuestro marco sobre la extrapolación. Esto es particularmente importante a la luz del argumento, presentado por algunos críticos, de que una consecuencia no deseada de la revolución de la credibilidad ha sido empujar a los académicos a centrarse en los datos de entornos muy localizados, lo que limita nuestra capacidad de aprender sobre principios más generales.

Vincular la teoría y lo empírico. Se centra en un objetivo de investigación en la realidad u objeto de estudio para abreviar. Un objeto es parte del universo real sobre el que estamos tratando de aprender, ya sea con teoría o empíricamente. Un modelo toma como objeto las leyes de la naturaleza, al igual que un estudio empírico que relaciona los resultados sintéticos de una teoría. Del mismo modo, la investigación teórica o empírica sobre los eventos de procesos tienen como objetivo la predicción en contextos reales.

Los modelos teóricos y los diseños de investigación empírica, por supuesto, son distintos al objeto en lo real. Pero cada uno, a su manera, intenta representar el objeto. Así que nuestra visión desde las alturas del bosque, debe aclarar un poco sobre lo que significa para los modelos teóricos y los diseños de investigación empírica representar objetos del mundo real. Una vez que podamos ver más claramente cómo los modelos y los diseños experimentales se relacionan con el mundo, también obtendremos algo de claridad sobre cómo se relacionan entre sí.

Comencemos con los modelos y la idea de similitud¹⁵. La investigación teórica implica tratar de explicar un objeto del mundo real mediante la representación de algunas características de ese objeto. Para ello, el modelo característico de ese objeto, en un sentido debe ser aclarado en lo similar al objeto real. La palabra modelo recuerda a los modelos formales de elección racional, y esos son un ejemplo paradigmático de la idea. Pero las nociones de representación y similitud se aplican más ampliamente. Se pueden aplicar a otras formalizaciones, incluidos los modelos computacionales y los esquemas conceptuales para formalizar relaciones causales desde una experiencia sintética. Y se aplican a una variedad de enfoques para teorizar informalmente, por ejemplo, el modelo de opinión pública.

Los modelos son siempre simplificaciones, ignorando gran parte de lo que está sucediendo en el objeto. A menudo hacen suposiciones que son demostrables parcialmente falsas como descripciones del mundo. Entonces, ¿cómo podrían ser útiles para la explicación teórica? La respuesta viene de desarrollar la noción de similitud, que da sustento a la práctica. Un modelo está destinado a ser similar a el objeto de una manera muy particular. Ciertos actores, situaciones y mecanismos análogos en el destino. Las llamamos las características representacionales del modelo. No todo en el modelo está destinado a ser representacional. Cualquier modelo teórico tiene características auxiliares que no son descriptivas del objeto, pero ayudan con la manejabilidad o simplificación necesaria para que ese conocimiento sea ingeniería. Las notificaciones sobre las entidades de representación definen las condiciones de ámbito del modelo.

El análisis de un modelo produce implicaciones sobre las relaciones entre los objetos del modelo. Algunas de esas implicaciones dependen de las entidades de representación, pero son robustas para los cambios en las entidades auxiliares. Esas implicaciones son adecuadas para evaluar la similitud del modelo y el objeto. A esto lo llamamos las implicaciones pertinentes y necesarias para el progreso de una tecnología. La idea clave es que, si la similitud es cierta, entonces las implicaciones pertinentes del modelo deben reflejarse en el objeto real. Si, por el contrario, la implicación teórica depende crucialmente de una característica auxiliar del modelo que no es plausible similar al mundo, entonces no hay razón para esperar ver la implicación reflejada en el mundo como una herramienta de predicción que desarrolla tecnología.

Confrontar la similitud teórica y empírica sobre un objeto del mundo real. No son solo los modelos teóricos los que están vinculados al mundo a través de la similitud. La investigación empírica implica un diseño experimental, también vinculado a un objeto del mundo real a través de una relación de similitud. El diseño de la investigación utiliza diversas medidas y procedimientos estadísticos para estimar alguna relación de interés. La similitud del diseño de la investigación con el mundo tiene dos aspectos. En primer lugar las medidas utilizadas deben representar significativamente las características relevantes del objetivo, es decir, deben tener lo que se llama validez de medición¹⁶. En segundo lugar, el diseño experimental proporciona buenas estimaciones de la relación de interés solo bajo ciertos supuestos. Esas suposiciones deberían describir plausiblemente el objetivo. Si ambos aspectos de la relación de similitud son ciertos, entonces las estimaciones generadas por el diseño reflejan la relación especificada entre las características relevantes del mundo.

Podemos vincular la teoría y los empíricos entre sí cuando hay cierta superposición en el conjunto de objetos a los que apuntan. En tal caso, podemos aprender acerca de la similitud comparando las implicaciones generadas por los modelos teóricos con las estimaciones generadas por los diseñadores empíricos. Si un modelo teórico es similar a un objeto del mundo real, entonces las implicaciones pertinentes del modelo deben reflejarse en el objeto. Si un diseño experimental empírico es similar a un objeto del mundo real, luego las estimaciones que salen del diseño de la investigación reflejan las relaciones entre las características relevantes del objeto. Así que si ambas relaciones de similitud se mantienen, tenemos razones para esperar un acuerdo entre las implicaciones teóricas y las estimaciones empíricas. Y a la inversa, el desacuerdo nos da alguna razón para dudar de la combinación de esas afirmaciones de similitud.

Pero cuán informativo es el desacuerdo o acuerdo, depende de si las estimaciones empíricas son conmensurables, es decir, si la relación con el objeto que está siendo descrito por la teoría y la relación que está siendo descrito por los empíricos son la misma relación. Como tal, la conmensurabilidad dice que los hallazgos empíricos son máximamente informativos sobre la similitud cuando los diseños experimentales ofrecen estimaciones empíricas de relaciones análogas al modelo teórico.

Nuestro objetivo científico es acumular conocimiento de los mecanismos que explican fenómenos naturales importantes. Como tal, lo que realmente queremos aprender es la relación de similitud entre los modelos teóricos y el mundo real¹⁷.

0.5 Ninguna explicación científica es para siempre

Nuestro objetivo es acumular conocimiento de mecanismos que expliquen fenómenos biológicos importantes. Como tal, lo que realmente queremos aprender es la relación de similitud entre los modelos teóricos y el mundo real.

Este proceso comienza estudiando modelos que ya tenemos buenas razones para creer que podrían ser similares al objeto real de investigación. Si los efectos, causales y mecanismos incorporados no representan plausibilidad al análogo en el mundo y sus hallazgos empíricos, no harán mucho para convencernos de que tenemos una buena explicación de lo que está sucediendo.

Pero incluso con un modelo plausible en la mano, hay mucho trabajo por hacer. No observamos directamente los mecanismos que están funcionando en la maquinaria biológica, ni podemos probar directamente la similitud entre un modelo y el objeto real. En cambio, procedemos indirectamente, buscando evidencia a favor o en contra de la similitud al comparar las implicaciones de un modelo que incorpora el mecanismo que se cree están en funcionamiento con estimaciones empíricas de un nuevo diseño experimental con el mismo objetivo.

Aquí, nos enfrentamos a un desafío para aprender sobre los mecanismos teóricos de los sistemas biológicos. El acuerdo o desacuerdo entre una implicación teórica pertinente y un hallazgo empírico es informativo sobre la afirmación conjunta de que las dos condiciones: similitud más la conmensurabilidad se mantiene. Es decir, es informativo el modelo sobre la cadena de conexiones representadas en cada eslabón, débil o fuerte en su causalidad. Si una implicación teórica y un hallazgo empírico no están de acuerdo, tal vez significa que las medidas del diseño de experimentación no son válidas o sus suposiciones no son plausibles, o tal vez significa que las simplificaciones teóricas y las estimaciones empíricas no son conmensurables.

Por lo tanto, cuánto aprendemos sobre qué mecanismos están en funcionamiento, depende críticamente de cuán fuertes sean nuestras razones para creer que el diseño experimental responde similar al objeto y que las implicaciones teóricas y las estimaciones empíricas son conmensurables. Si el diseño experimental utiliza buenas medidas y suposiciones plausibles para estimar una cantidad altamente conmensurable, entonces el acuerdo o el desacuerdo es muy informativo sobre los mecanismos causales. Pero a medida que disminuye la calidad de las medidas, la plausibilidad de los supuestos o la conmensurabilidad de las estimaciones empíricas y las implicaciones teóricas, también lo hace cuánto podemos aprender sobre los mecanismos incorporados en el modelo biológico.

Esta es la razón por la que la revolución de la credibilidad dentro de la ciencia de datos es tan importante para formar un vínculo productivo entre teoría y empírica. Se centra el diseño experimental biológico en hacer creíble las predicciones y la afirmación de haber estimado el tipo de cantidad que es conmensurable con implicaciones teóricas en particular, relaciones de igualdad de todo lo demás.

Dicho esto, nuestro marco de pensamiento sugiere que no debemos dejar que grandes volúmenes de datos o simples herramientas estadísticas sean enemigos de la buena determinación causal. La causalidad permite que aprendamos cosas de modelos imperfectos o diseños de investigación perfectibles. Pero cuando más estrechas podamos hacer las conexiones causales, la similitud del modelo o diseño experimental con el objeto y la conmensurabilidad de las implicaciones teóricas y estimaciones empíricas, más aprenderemos. Por lo tanto, el progreso científico de una agenda de investigación puede tomar la forma de fortalecer las razones para creer en cualquiera de estas afirmaciones.

Ninguna cantidad de observaciones por debajo de esta altura intelectual, sería suficiente para resolver la cuestión de hacer de la biología una ciencia de predicciones. Desafortunadamente, a la juventud universitaria es común que su educación esté al margen del modelado y su evaluación, están fuera del radar de sus planes de estudio; y más grave, su capacidad de pensar al escribir y leer (disertación) es débil, inferimos que cortar y pegar texto es el mayor signo de esta tragedia.

De esta manera, una línea de investigación avanza a través de un ir y venir entre la teoría y la empírica. En una dirección los modelos teóricos pueden guiar al científico a los datos. Sin un modelo teórico implica alguna relación, entonces puede proceder buscando un diseño de investigación creíble para estimar una cantidad conmensurable en el fin de evaluar la similitud del modelo teórico con el objeto. En otra dirección, los hallazgos empíricos pueden guiar la teorización. Si un diseño de investigación ha estimado alguna relación en los datos, puede proceder buscando un modelo plausiblemente similar con una implicación conmensurable para explicar el hallazgo empírico.

0.6 Modelado matemático de sistemas biológicos

La biología de sistemas adopta un enfoque interdisciplinario para el estudio sistemático de interacciones complejas en sistemas biológicos. Este enfoque busca descifrar los comportamientos emergentes de sistemas complejos en lugar de centrarse solo en sus propiedades constituyentes. Con el objetivo de comprender las interacciones dinámicas entre los componentes de una célula y entre las células, así como su interacción con el medio ambiente, la biología de sistemas es un enfoque mediante el cual se abordan cuestiones biomédicas mediante la integración de experimentos en ciclos iterativos con modelos matemáticos, simulación y teoría. El modelado no es el objetivo final, sino una herramienta para aumentar la comprensión del sistema, desarrollar experimentos más dirigidos y, finalmente, permitir predicciones. Los modelos matemáticos tienen la ventaja de ser cuantitativos e interactivos en lugar de únicamente descriptivos. El proceso mediante el cual se formulan los modelos, que puede incluir la representación de efectos genéticos, epigenéticos, celulares y tisulares en las diversas escalas físicas y temporales durante la tumorigénesis, ayuda a articular hipótesis y, por lo tanto, apoya el diseño de experimentos apropiados para probarlas¹⁸. Las definiciones más populares de biología de sistemas se refieren a dinámicas, mecanismos, principios y comportamientos. La complejidad de los sistemas y/o funciones biológicos surge de la interacción de una observación de fenómenos y componentes espacio-temporales no lineales. El hecho de que la mayoría de los procesos celulares, como el control del ciclo celular, la diferenciación celular y la apoptosis, sean inherentemente dinámicos, destaca la necesidad de integrar modelos matemáticos en las ciencias de la vida y la investigación clínica. Un enfoque de biología de sistemas puede ayudar a identificar y analizar los principios, leyes y mecanismos subyacentes al comportamiento de los

sistemas biológicos. En biología de sistemas, los argumentos surgen sobre las predicciones y la validez de las teorías, los métodos de recopilación de datos y la interpretación de conjuntos de datos experimentales. El papel del modelado matemático dentro del campo de la biología de sistemas, mediando la interpretación de datos experimentales y ayudando a la formulación de hipótesis. Una garantía es la justificación que explica la relación de los datos con la hipótesis (afirmación). A menudo, las garantías se basan en supuestos contextuales que solo se reconocen tácitamente. Los calificadores expresan los límites de la validez del reclamo. Los argumentos surgen cuando se intenta refutar, ya sea atacando la validez de los datos o la validez de la orden judicial. El modelado matemático encaja en la filosofía de argumentación de Toulmin¹⁹. La principal razón por la que se necesita el modelado en las ciencias de la vida es la complejidad de los sistemas naturales. El número de componentes realmente no juega un papel particular en esto. Tener muchas moléculas o células interactuando no es un problema como tal. Sin embargo, el número de diferentes tipos de componentes constituye un desafío para la teoría de los sistemas no lineales, que hasta el día de hoy se practica solo para un puñado de variables del sistema.

El hecho de que una molécula y su forma modificada (digamos por fosforilación) requieran ya dos variables de sistemas en un modelo de ecuaciones diferenciales, muestra que los sistemas con diez o veinte especies moleculares pueden volverse bastante difíciles de manejar. También se deben agregar las dificultades para dar cuenta de los fenómenos espaciales. Si bien, dominan los modelos basados en ecuaciones diferenciales ordinarias, en la actualidad no es práctico formular modelos de ecuaciones diferenciales parciales e identificar sus parámetros a partir de datos experimentales. La no linealidad en las interacciones y el comportamiento resultante es otro elemento importante de complejidad, un obstáculo y una fuente de sorpresa. Sin embargo, lo más importante es el hecho de que un sistema natural cambia, se adapta y evoluciona constantemente, lo que dificulta el uso de enfoques que asuman estacionariedad e invariancias en el tiempo. Un sistema vivo, como una célula, un tejido, un órgano u organismo, actúa y reacciona; responde y modifica su entorno. Las células, como los organismos, se someten a un proceso continuo de interacción y cambios mutuos. Un sistema vivo está limitado por su entorno, pero también cambia su entorno. Esta autorreferencia y los fenómenos que surgen posteriormente son la verdadera causa de problemas para el modelador.

Si bien el modelado matemático (especialmente de sistemas dinámicos) es un elemento central de la biología de sistemas, el campo de la bioinformática tiene más que ver con el análisis de datos e información, ya sea directamente de experimentos o de bases de datos. Ambas áreas son complementarias y dependen la una de la otra. Por ejemplo, para simular una vía de señalización, la construcción del modelo se beneficia del conocimiento sobre las propiedades estructurales, por ejemplo, los sitios de fosforilación, de una proteína. Esto se puede encontrar en bases de datos utilizando métodos y herramientas bioinformáticas. Para muchos sistemas, no podemos desarrollar modelos mecanicistas detallados debido a la falta de datos cuantitativos del curso del tiempo. No obstante, con la ayuda de la bioinformática se pueden analizar bases de datos para ayudar a formular y verificar hipótesis sobre redes a un nivel superior de abstracción. Se deben reconocer los desafíos del modelado mecanicista y el papel complementario que desempeñan los métodos bioinformáticos en el tratamiento de la incertidumbre que surge de la complejidad de los sistemas en consideración.

Hay dos paradigmas dominantes que se utilizan en el modelado matemático de las redes de reacciones bioquímicas (pathways) en la biología de sistemas: el enfoque determinista, que utiliza simulaciones numéricas de ecuaciones diferenciales ordinarias no lineales (incluidos los modelos de acción de masas, ley de potencia o de Michaelis-Menten), y el enfoque estocástico basado en una ecuación maestra y simulaciones estocásticas. El modelado estocástico tiene una larga tradición en física, muestra cómo se puede formular un modelo estocástico que comprendan tanto las leyes deterministas como las fluctuaciones sobre ellas. A veces, estos modelos se denominan modelos "mesoscópicos". Considerando un sistema de puntos de masa que interactúan, las fluctuaciones en los sistemas de no equilibrio no surgen de una distribución de probabilidad del microestado inicial, sino que son generadas continuamente por las ecuaciones de movimiento de las moléculas. Si bien, los modelos estocásticos mesoscópicos son conceptos teóricos atractivos, en un contexto práctico en el que dicho modelo (no lineal) y los valores de sus parámetros tendrían que extraerse de datos experimentales, nos enfrentamos a varios problemas (que son en parte una razón para el amplio uso de las ecuaciones diferenciales).

Podemos ilustrar las nociones de microscópico, mesoscópico y macroscópico en el contexto de la biología celular considerando la expresión génica, el proceso por el cual

la información del genoma se transcribe primero en ARN antes de traducirse en proteínas. Estas dos etapas involucran dos niveles, la transcripción de un gen es microscópica en comparación con las fluctuaciones en la concentración de la proteína para la cual el gen codifica la información. Mientras que para el inicio de la transcripción, digamos a través de la unión de factores de transcripción, un modelo estocástico discreto puede ser apropiado, los cambios en las concentraciones de las proteínas involucradas en la función de una sola célula (p. Ej., Ciclo celular) pueden ser por otro lado descrito macroscópicamente por ecuaciones diferenciales ordinarias. Sin embargo, eso solo es válido si los efectos de eventos aleatorios discretos en el nivel de transcripción no se propagan al nivel de traducción. Cuando eso suceda, el modelo continuo que describe los cambios en las concentraciones de proteínas también debe ser estocástico y tomará la forma de la llamada "ecuación de Langevin".

En muchas situaciones, las fluctuaciones aleatorias son lo suficientemente pequeñas como para ignorarlas, lo que permite que las ecuaciones macroscópicas predigan el comportamiento de un sistema con gran precisión. Las células, sin embargo, son sistemas abiertos, donde el entorno puede forzarlas a un estado de no equilibrio estacionario en el que la dinámica del sistema se bifurca, la dirección tomada depende de las fluctuaciones específicas que ocurren. Tenga en cuenta que, por lo tanto, la aleatoriedad de las fluctuaciones (que solo se puede describir en términos de probabilidades) influye de manera más crítica en el comportamiento del sistema de ecuaciones macroscópicas en puntos de bifurcación específicos, mientras que otras áreas del espacio de estados pueden aproximarse perfectamente mediante ecuaciones macroscópicas. El ruido intrínseco de las fluctuaciones térmicas o el control transcripcional podría determinar cómo el sistema a nivel macroscópico atraviesa una bifurcación. Al observar una población de células genéticamente idénticas en un entorno homogéneo, esto conduce a la variabilidad de los estados celulares que bien pueden ser explotados por el sistema biológico²⁰. El contexto obvio en el que la aleatoriedad tiene una función está generando diversidad en la evolución. Al observar un solo gen en una sola célula, el inicio de la transcripción en su sitio promotor es impulsado por la asociación y disociación de un número muy pequeño de moléculas. Este número muy bajo de copias de moléculas tiene dos consecuencias: el tiempo de los eventos de reacción se puede describir solo en términos de probabilidades, y los cambios en el número de moléculas son discretos, sin una tendencia obvia que pueda aproximarse con una ecuación diferencial²¹. Sin embargo, la expresión de un gen

cumple una función celular como el crecimiento, la diferenciación y la apoptosis. Por ejemplo, en respuesta a estímulos externos, la célula puede producir grandes cantidades de una proteína. Esta respuesta, medida como un cambio de concentración aparentemente suave/monótono, se describe adecuadamente mediante ecuaciones diferenciales. Por tanto, se ignoran las pequeñas fluctuaciones en torno a una tendencia/media obvia. En este nivel, estamos apuntando a una descripción de una vía que actúa como interruptor, filtro, oscilador o amplificador, estudiando el comportamiento de la red en términos de su robustez, capacidad de respuesta y sensibilidad del modelo a cambios en los parámetros, transiciones entre estados estacionarios, y bifurcaciones. Una suposición habitual en tales modelos de ecuaciones de tasa es que los parámetros (coeficientes de tasa) son constantes. Dado que estos parámetros están implícitamente vinculados a variables ambientales, como la temperatura, el nivel de pH o el balance hídrico, las fluctuaciones en estos se consideran insignificantes. El arte de modelar es entonces decidir en el contexto dado, qué enfoque de modelado o combinación de enfoques es más apropiado.

0.7 Punto de partida

La biología es el estudio de los seres vivos; La mecánica es el estudio de los movimientos y las cargas aplicadas que los provocan. La biomecánica se puede definir, por tanto, como el estudio de los movimientos que experimentan los seres vivos en respuesta a las cargas aplicadas. Aquí, sin embargo, consideramos que la biomecánica es el desarrollo, extensión y aplicación de la mecánica con el propósito de comprender mejor la influencia de las cargas aplicadas en la estructura, propiedades y función de los seres vivos y las estructuras con las que interactúan. Por tanto, el dominio de la biomecánica es muy amplio. Incluye, entre muchas otras cosas, el estudio de los efectos de las cargas de viento o la gravedad en el crecimiento de las plantas, las propiedades mecánicas de los alimentos, el vuelo de las aves, las propiedades reductoras del arrastre de la piel de los delfines y el rendimiento atlético humano. Además, la biomecánica aborda muchos problemas de salud, así como enfermedades, lesiones y su tratamiento tanto en humanos como en animales. Esta será nuestra principal motivación aquí; por lo tanto, es fácil ver que la biomecánica es fundamental para el campo de rápido crecimiento de la ingeniería biomédica. No es posible identificar a un verdadero "padre de la biomecánica", pero muchos apuntan a Leonardo da Vinci (1452-1519) o Galileo Galilei (1564-1642). Entre muchas otras cosas, da Vinci estaba interesado en un medio

por el cual el hombre pudiera volar y, con este fin, estudió la mecánica del vuelo de los pájaros. El intento de la humanidad de basar el diseño de sistemas de ingeniería en la forma en que la naturaleza hace algo (por ejemplo, la estructura de panal dentro de una colmena o el sistema de radar de un murciélago) se llama biónica, que sigue siendo un área muy importante dentro de la biomecánica. A diferencia de da Vinci, Galileo estaba interesado en la fuerza intrínseca de los huesos y, en particular, su relación con el diseño estructural de los huesos. Basado en un análisis preliminar, sugirió que los huesos son huecos, por esto mejora la relación resistencia-peso. Claramente, entonces, la biomecánica se enfoca tanto en el diseño como en el análisis, cada uno de los cuales es fundamental para la ingeniería. Saltando hacia el final del siglo XIX, Wilhelm Roux propuso la idea de un "mecanismo de autorregulación cuantitativo" que da como resultado la adaptación funcional de tejidos, órganos y organismos, una idea que era consistente con el concepto de una organización mediada por estrés de la microestructura del hueso que fue presentada por Julius Wolff en 1884. Brevemente, Wolff sugirió que la estructura fina dentro de los huesos (es decir, trabéculas orientadas) está gobernada por líneas de tensión que resultan de las cargas aplicadas. Aunque su análisis no fue correcto, la idea básica fue sumamente importante. De hecho, volveremos muchas veces a esta observación de que las cargas mecánicas controlan la estructura y función de los tejidos, lo que ha dado lugar a un área de investigación muy importante llamada mecanobiología. Muchos otros sabios estaban interesados en aplicaciones biomecánicas. Entre ellos se encuentran R. Hooke (1635-1703), L. Euler (1707-1783), T. Young (1773-1829), J.L.M. Poiseuille (1799–1869) y H. von Helmholtz (1821–1894). A pesar del calibre de los científicos que han buscado respuestas en biomecánica a lo largo de los siglos, nuestro campo no se desarrolló realmente hasta mediados de la década de 1960. Aunque los historiadores probablemente discutirán sobre las razones de esto, se sugiere aquí que cinco desarrollos casi simultáneos proporcionaron tanto una mayor motivación como una mayor capacidad en biomecánica. Recuerde que la década de 1960 fue la década de la búsqueda de la Luna por parte de la humanidad. Ante la pregunta, "¿Cómo responderá el hombre a las cargas alteradas asociadas con los viajes espaciales, incluida una carga gravitacional reducida en la Luna?", La medicina clínica no pudo proporcionar las respuestas, ya que se basa principalmente en observaciones. Por lo tanto, se necesitaba una ciencia predictiva que se centrara en cómo responde el cuerpo a las cargas mecánicas. Además, tenga en cuenta que gran parte de la biomecánica se ocupa de la respuesta de los tejidos blandos (es decir, tejidos distintos de los huesos y

los dientes). Se sabe desde hace mucho tiempo que los tejidos blandos exhiben comportamientos no lineales complejos que no podrían ser descritos por la mecánica clásica de los continuos desarrollados en los siglos XVIII y XIX. Más bien, la biomecánica tuvo que esperar el renacimiento posterior a la Segunda Guerra Mundial en la mecánica del continuo (1948-1965) a través del cual las teorías no lineales lograron una base más completa y racional. Durante este mismo período, de los años 50 a los 60, los desarrollos tecnológicos dieron lugar a la computadora digital. Las computadoras son esenciales en biomecánica para resolver muchos problemas importantes pero complejos de límites y valores iniciales, para controlar experimentos complicados y para realizar análisis no lineales de los datos. Paralelamente al desarrollo de las computadoras, se produjo el avance de poderosos métodos numéricos de análisis, incluido el método de elementos finitos, que se introdujo en 1956 y se ha convertido en una herramienta estándar en el arsenal del biomecánico para atacar problemas básicos y aplicados. Finalmente, no es casualidad que la biomecánica surgiera en el momento en que nació la biología moderna, que se debió en gran parte a la identificación en la década de 1950 de la estructura básica de las proteínas (por L. Pauling) y el ADN (por J. Watson y F. Crick). En resumen, las décadas de 1950 y 1960 proporcionaron nuevas motivaciones importantes, así como avances teóricos, experimentales y tecnológicos que permitieron el surgimiento de la biomecánica. Esto es, por supuesto, solo una sinopsis de algunos de los desarrollos históricos esenciales. Se anima al lector interesado a indagar más en la historia de nuestro campo. Aunque la biomecánica abarca una amplia gama de temas, los propósitos de este libro son dos: primero, presentar conceptos fundamentales y resultados de la mecánica de sólidos y fluidos que se pueden aplicar a muchos problemas diferentes de importancia en biología y medicina y, segundo, ilustrar algunas de las muchas aplicaciones posibles al centrarse en la mecánica de la salud, las enfermedades y las lesiones humanas. Por lo tanto, para motivar aún más nuestro estudio, revisemos brevemente algunos de los muchos casos en los que la biomecánica puede y debe contribuir al avance de la atención médica.

0.8 Introducción a los mecanismos

En muchos campos de la ciencia, lo que se considera una explicación satisfactoria requiere proporcionar una descripción de un mecanismo. Por tanto, no es sorprendente que gran parte de la práctica de la ciencia pueda entenderse en términos

del descubrimiento y descripción de mecanismos. Nuestro objetivo es esbozar un enfoque mecanicista para analizar la biología molecular que se basa en detalles de la práctica científica, un enfoque que bien puede aplicarse a otros campos científicos. Los mecanismos se han invocado en muchas ocasiones y lugares en la filosofía y la ciencia. Una búsqueda por palabra clave "mechanisms" para 2021 en títulos y resúmenes de Nature encontró 288,246 resultados. Sin embargo, en nuestra opinión, no existe un análisis adecuado de qué son los mecanismos y cómo funcionan en la ciencia. Comenzamos con un análisis dualista del concepto de mecanismo en términos tanto de las entidades como de las actividades que los componen. Estos argumentos no se desarrollan en detalle, pero deberían ser suficientes para mostrar cómo pensar en los mecanismos proporciona un enfoque distintivo a muchos problemas en la ciencia. Sin embargo, hacemos algunas advertencias rápidamente. Primero, usamos "mecanismo" porque la palabra se usa comúnmente en ciencia. Pero como detallaremos con mayor precisión, no se debe pensar en los mecanismos como sistemas exclusivamente mecánicos (push-pull). Lo que cuenta como un mecanismo en la ciencia se ha desarrollado con el tiempo y presumiblemente seguirá haciéndolo. En segundo lugar, limitaremos nuestra atención a los mecanismos de la biología. No afirmamos que todos los científicos busquen mecanismos o que todas las explicaciones sean descripciones de mecanismos. Sospechamos que este análisis es aplicable a muchas otras ciencias, y tal vez incluso a mecanismos cognitivos o sociales, pero lo dejamos como una cuestión abierta. Por último, muchos de nuestros puntos se enuncian solo de manera provocativa y breve. Creemos que hay argumentos completos para estos puntos, pero detallarlos aquí oscurecería la visión general.

0.8.1 Mecanismos

Se buscan mecanismos para explicar cómo se produce un fenómeno o cómo funciona algún proceso significativo. Específicamente:

Los mecanismos son entidades y actividades organizadas de manera que produzcan cambios regulares desde el inicio o la configuración hasta el final o las condiciones de terminación.

Por ejemplo, en el mecanismo de la neurotransmisión química, una neurona presináptica transmite una señal a una neurona postsináptica liberando moléculas de

neurotransmisores que se difunden a través de la hendidura sináptica, se unen a los receptores y despolarizan la célula postsináptica. En el mecanismo de replicación del ADN, la doble hélice del ADN se desenrolla, exponiendo bases ligeramente cargadas a las que se unen bases complementarias, produciendo, después de varias etapas más, dos hélices duplicadas. Las descripciones de los mecanismos muestran cómo las condiciones de terminación son producidas por las condiciones de configuración y las etapas intermedias. Dar una descripción de un mecanismo de un fenómeno es explicar ese fenómeno, es decir, explicar cómo se produjo. Los mecanismos se componen de entidades (con sus propiedades) y actividades. Las actividades son las productoras de cambio. Las entidades son las cosas que participan en actividades. Las actividades generalmente requieren que las entidades tengan tipos específicos de propiedades. El neurotransmisor y el receptor, dos entidades, se unen, una actividad, en virtud de sus propiedades estructurales y distribuciones de carga. Una base de ADN y un enlace de hidrógeno de base complementaria debido a sus estructuras geométricas y cargas débiles. La organización de estas entidades y actividades determina las formas en que producen el fenómeno. Las entidades a menudo deben estar ubicadas, estructuradas y orientadas de manera adecuada, y las actividades en las que se involucran deben tener un orden, una tasa y una duración temporales. Por ejemplo, dos neuronas deben estar espacialmente próximas para la difusión del neurotransmisor. Los mecanismos son regulares en el sentido de que funcionan siempre o en su mayor parte de la misma manera en las mismas condiciones. La regularidad se exhibe en la forma típica en que el mecanismo corre de principio a fin; lo que lo hace regular es la continuidad productiva entre etapas. Las descripciones completas de los mecanismos exhiben una continuidad productiva sin brechas desde la configuración hasta las condiciones de terminación. Las continuidades productivas son las que hacen inteligibles las conexiones entre etapas. Si un mecanismo está representado esquemáticamente por $A \rightarrow B \rightarrow C$, entonces la continuidad radica en las flechas y su explicación está en términos de las actividades que representan las flechas. Una flecha faltante, es decir, la incapacidad de especificar una actividad, deja un vacío explicativo en la continuidad productiva del mecanismo. No somos los únicos que pensamos que el concepto de "mecanismo" es fundamental para una comprensión filosófica adecuada de las ciencias biológicas. Otros han defendido la importancia de los mecanismos en biología²². Wimsatt, por ejemplo, dice que, "al menos en biología, la mayoría de los científicos ven su trabajo cómo explicar tipos de fenómenos mediante el descubrimiento de mecanismos²³...". Schaffner a menudo señala la importancia de los mecanismos en

biología y medicina, pero argumenta, siguiendo a Mackie, que hablar de mecanismos causales depende de un discurso previo y más fundamental sobre las "leyes del trabajo"²⁴. En otra parte, Schaffner afirma que "mecanismo", tal como lo utilizan Wimsatt y otros, es un "término no analizado" que desea evitar. La noción de "mecanismo" ha sido analizado típicamente en términos de la descomposición de sistemas en sus partes e interacciones²⁵. Siguiendo esta tradición "interaccionista", Glennan define un mecanismo de la siguiente manera: un mecanismo subyacente a un comportamiento es un sistema complejo que produce ese comportamiento... la interacción de varias partes de acuerdo con leyes causales directas²⁶.

Afirma que todas las leyes causales se explican proporcionando un mecanismo de nivel inferior hasta que uno toca fondo en las leyes fundamentales no causales de la física. Consideramos que la confianza de Glennan en el concepto de "ley" es problemática porque, en nuestros ejemplos, rara vez hay "leyes causales directas" para caracterizar cómo operan las actividades. Más importante aún, nos parece que la confianza del interaccionista en las leyes y las interacciones deja fuera la naturaleza productiva de las actividades. Nuestra forma de pensar enfatiza las actividades en los mecanismos. El término "actividad" trae consigo las connotaciones apropiadas de su uso estándar; sin embargo, tiene la intención de ser un término técnico. Por lo general, una actividad se designa mediante un verbo o una forma verbal (participios, gerundios, etc.). Las actividades son las productoras de cambio. Son constitutivos de las transformaciones que producen nuevos estados de cosas o nuevos productos. La referencia a las actividades están motivada por preocupaciones ónticas, descriptivas y epistemológicas. Justificamos esta ruptura con la parsimonia, este dualismo de entidades y actividades, por referencia a estas necesidades filosóficas.

0.8.2 Estado óntico de los mecanismos

Tanto las actividades como las entidades deben estar incluidas en una adecuada cuenta óntica de mecanismos. Nuestro análisis del concepto de mecanismo es explícitamente dualista. Estamos intentando captar las intuiciones filosóficas saludables que subyacen a las ontologías sustantivistas y de proceso. Las sustantivistas limitan su atención a entidades y propiedades, asumen que es posible reducir la narrativa de sus actividades a hablar de propiedades y sus transiciones. Sustantivistas hablan así de entidades con capacidades²⁷ o disposiciones para actuar. Sin embargo, para identificar la capacidad de

una entidad, primero se deben identificar las actividades en las que participa esa entidad. No se sabe que la aspirina tiene la capacidad de aliviar un dolor de cabeza a menos que se sepa que la aspirina produce alivio del dolor de cabeza. Los sustantivistas también hablan sobre interacciones de entidades o sus transiciones de estado. Creemos que las transiciones de estado deben describirse de manera más completa en términos de las actividades de las entidades y cómo esas actividades producen cambios que constituyen la siguiente etapa. Lo mismo ocurre con el habla de interacciones, que enfatiza las intersecciones espacio-temporales y los cambios en las propiedades sin caracterizar la productividad por la cual esos cambios se efectúan en esas intersecciones. Los sustantivistas enfocan apropiadamente la atención sobre las entidades y propiedades en los mecanismos, por ejemplo, el neurotransmisor, el receptor y sus configuraciones de carga o bases de ADN y sus débiles polaridades. Son las entidades las que realizan actividades, y lo hacen en virtud de algunas de sus propiedades. Esta es la razón por la que son las relaciones de relevancia estadística²⁸ entre las propiedades de las entidades en un momento y las propiedades de las entidades en otro (o generalizaciones que establecen relaciones "entrada-salida" y cambios de estado) son útiles para describir mecanismos. Sin embargo, es artificial y empobrecido describir los mecanismos únicamente en términos de entidades, propiedades, interacciones, entradas-salidas y cambios de estado a lo largo del tiempo. Los mecanismos hacen cosas. Son activos y, por lo tanto, deberían describirse en términos de las actividades de sus entidades, no simplemente en términos de cambios en sus propiedades.

A diferencia de los sustantivistas, los ontólogos de procesos cosifican las actividades e intentan reducir las entidades a procesos²⁹. Si bien la ontología de procesos reconoce la importancia de los procesos activos al tomarlos como unidades ontológicas fundamentales, su programa para la reducción de entidades es, en el mejor de los casos, problemático. Hasta donde sabemos, no hay actividades en neurobiología, biología molecular y nanobiología que no sean actividades de entidades. No obstante, los ontólogos de procesos destacan apropiadamente la importancia de los tipos activos de cambio. Hay tipos de cambios al igual que existen tipos de entidades. Estos diferentes tipos son reconocidos por la ciencia y son básicos para la forma en que funcionan las cosas. Las actividades se identifican e individualizan de la misma manera que las entidades. Tradicionalmente, uno identifica e individualiza entidades en términos de sus propiedades y ubicación espacio-temporal. Asimismo, las actividades

pueden identificarse e individualizarse por su ubicación espacio-temporal. También pueden ser individualizadas por su tasa, duración, tipos de entidades y tipos de propiedades que participan en ellas (causales). Las condiciones de individuación más específicas pueden incluir su modo de operación (p. Ej., acción de contacto *versus* atracción a distancia), direccionalidad (p. Ej., lineal *versus* en ángulo recto), polaridad (atracción *versus* atracción y repulsión), requerimientos de energía (p. Ej., cuánta energía se requiere para formar o romper un enlace químico), y el rango de actividad (por ejemplo, las fuerzas electromagnéticas tienen una influencia más amplia que las fuerzas fuertes y débiles en el núcleo). A menudo, generalizaciones o las leyes son enunciados cuyos predicados se refieren a las entidades y propiedades que son importantes para la individuación de las actividades. Los mecanismos son identificados e individualizados por las actividades y entidades que los constituyen, por sus condiciones de inicio y finalización, y por sus **roles funcionales**. Las funciones son los roles que desempeñan las entidades y actividades en un mecanismo. Ver una actividad como una **función** es verla como un componente de algún mecanismo, es decir, verla en un contexto que se considera importante, vital o significativo de alguna otra manera. Es común hablar de funciones como propiedades "poseídas por" entidades, como cuando se dice que el corazón "tiene" la función de bombear sangre o el canal "tiene" la función de bloquear el flujo de sodio. Esta forma de hablar refuerza la tendencia sustantivista contra la que hemos estado argumentando. Las funciones, más bien, deben entenderse en términos de las actividades en virtud de las cuáles las entidades contribuyen al funcionamiento de un mecanismo. Es más apropiado decir que la función del corazón es bombear sangre y así entregar (con la ayuda del resto del sistema circulatorio) oxígeno y nutrientes al resto del cuerpo. Del mismo modo, una función de los canales de sodio es la puerta de la corriente de sodio en la producción de potenciales de acción. En la medida en que la actividad de un mecanismo en su conjunto contribuya a algo en un contexto que se considere previamente importante, vital o de otra manera significativa, esa actividad también puede considerarse como la (o una) función del mecanismo como en su conjunto³⁰. Las entidades y un subconjunto específico de sus propiedades determinan las actividades en las que pueden participar. Por el contrario, las actividades determinan qué tipos de entidades (y qué propiedades de esas entidades) son capaces de ser la base de tales actos. Dicho de otra manera, las entidades que tienen ciertos tipos de propiedades son necesarias para la posibilidad de actuar de ciertas formas específicas, y ciertos tipos de actividades solo son posibles cuando hay entidades que tienen ciertos tipos de propiedades. Las entidades y

actividades son correlativas. Son interdependientes. Una descripción ópticamente adecuada de un mecanismo incluye ambos.

0.8.3 Actividades y causas

Las actividades son tipos de causas. Términos como "causar" e "interactuar" son términos abstractos que deben especificarse con un tipo de actividad y, a menudo, así se especifican en el discurso científico típico. Bugajak (2011) señaló que la palabra "causar" en sí misma es muy general y solo adquiere significado cuando se completa con otros verbos causales más específicos, por ejemplo, raspar, empujar, secar, transportar, comer, quemar, derribar³¹. Una entidad actúa como causa cuando se dedica a una actividad productiva. Esto significa que se puede decir que los objetos simplificados, o incluso los tipos naturales, son causas solo en un sentido derivado. No es la penicilina lo que hace que desaparezca la neumonía, sino lo que hace la penicilina.

El intento de Mackie de analizar la necesidad de la causalidad en términos de las leyes del trabajo es similar a nuestro análisis en muchos aspectos³². Esto enfatiza que las leyes del trabajo deben descubrirse empíricamente y no se encuentran *a priori*. También afirma que los contrafactuales están respaldados por la evidencia inductiva de que tales procesos básicos están en funcionamiento. Sin embargo, quiere analizar la causalidad en términos de continuidad cualitativa o estructural de procesos, y más vagamente en términos de "fluir desde" o "extruir"³³. No está claro cómo aplicar tales conceptos en nuestros casos biológicos. Pero tal vez se están tratando de usarlos para referirse a lo que llamamos "actividades" y para captar lo que entendemos por "productividad".

Nuestro énfasis en los mecanismos es compatible, de alguna manera, con la filosofía mecánica de Salmon, ya que los mecanismos se encuentran en el corazón de la filosofía mecánica³⁴. Los mecanismos, para Salmon, se componen de procesos (cosas que exhiben consistencia de características a lo largo del tiempo) e interacciones (intersecciones espacio-temporales que involucran cambios persistentes en esos procesos). Es apropiado comparar nuestra narrativa de actividades con la de interacciones de Salmon. Salmon identifica interacciones en términos de marcas transmitidas y relaciones de relevancia estadística y, más recientemente, en términos de intercambio de cantidades conservadas³⁵. Aunque reconocemos la posibilidad de que el análisis de Salmon pueda ser todo lo que hay para ciertos tipos fundamentales de

interacciones en física, su análisis guarda silencio en cuanto al carácter de la productividad en las actividades investigadas por muchas otras ciencias. La mera narrativa sobre la transmisión de una marca o el intercambio de una cantidad conservada no agota lo que estos científicos saben sobre las actividades productivas y sobre cómo las actividades efectúan cambios regulares en los mecanismos. Por ejemplo, gran parte de lo que hacen los neurobiólogos y biólogos moleculares debe verse como un esfuerzo por comprender estos diversos tipos de producción y las formas en que funcionan.

0.8.4 Actividades y leyes

La noción tradicional de una ley universal de la naturaleza tiene pocas aplicaciones, si es que tiene alguna, en neurobiología o biología molecular. A veces, las leyes pueden describir la **regularidad de las actividades**. A veces no pueden. Por ejemplo, la ley de Ohm se usa para describir aspectos de las actividades en los mecanismos de neurotransmisión. No existe ninguna ley que describa las regularidades de la unión de proteínas a regiones del ADN. No obstante, la noción de actividad conlleva algunos de los rasgos característicos asociados a las leyes. Las leyes se toman como regularidades determinadas. Describen algo que actúa de la misma manera bajo las mismas condiciones, es decir, misma causa, mismo efecto³⁶. Es la misma forma en que hablamos de los mecanismos y sus actividades. Un mecanismo es la serie de actividades de las entidades que provocan las condiciones de terminación o terminación de forma regular. Estas regularidades no son accidentales y apoyan contrafactuales en la medida en que describen actividades. Por ejemplo, si se cambiara esta base única en el ADN y el mecanismo de síntesis de proteínas funcionara como de costumbre, entonces la proteína producida tendría un sitio activo que se une con más fuerza. Esta contrafactual justifica hablar de mecanismos y sus actividades con algún tipo de necesidad. No se realiza ningún trabajo filosófico postulando algo más, una ley, que suscribe la productividad de las actividades. En resumen, somos dualistas: tanto las entidades como las actividades constituyen mecanismos. No hay actividades sin entidades y las entidades no hacen nada sin actividades. Hemos argumentado a favor de la adecuación óptica de este dualismo al mostrar que puede captar los conocimientos tanto de los sustantivistas como de los ontólogos de procesos, al mostrar cómo se necesitan actividades para especificar el término "causa" y mediante un análisis

de las actividades que muestra su regularidad y necesidad a veces caracterizada por leyes.

0.8.5 Actividades y leyes, jerarquías en bocetos

Esquemas de mecanismo y bocetos. Los mecanismos ocurren en jerarquías anidadas y las descripciones de los mecanismos en neurobiología y biología molecular son frecuentemente de varios niveles. Los niveles de estas jerarquías deben considerarse como jerarquías de parte-todo con la restricción adicional de que las entidades, propiedades y actividades de niveles inferiores son componentes de mecanismos que producen fenómenos de niveles superiores³⁷. Por ejemplo, la activación del canal de sodio es un componente del mecanismo de despolarización, que es un componente del mecanismo de neurotransmisión química, que es un componente de la mayoría de los mecanismos de nivel superior en el sistema nervioso central. Se pueden encontrar jerarquías similares en biología molecular. James Watson (1965) analiza los mecanismos para formar enlaces químicos fuertes y débiles, que son componentes de los mecanismos de replicación, transcripción y traducción de ADN y ARN, respectivamente, que son componentes de los mecanismos de numerosas actividades celulares³⁸.

0.8.6 Tocando fondo

Las descripciones jerárquicas anidadas de los mecanismos suelen tocar fondo en los mecanismos de nivel más bajo. Estos son los componentes que se aceptan como relativamente fundamentales o se consideran no problemáticos para los propósitos de un científico, grupo de investigación o campo determinado. Tocar fondo es relativo: los diferentes tipos de entidades y actividades son donde un campo dado se detiene al construir mecanismos. La explicación llega a su fin y la descripción de los mecanismos de nivel inferior sería irrelevante para sus intereses. Además, la formación científica a menudo se concentra en o alrededor de ciertos niveles de mecanismos. Los neurobiólogos con diferentes intereses teóricos o experimentales tocan fondo en distintos tipos de entidades y actividades. Algunos neurobiólogos están interesados principalmente en el comportamiento de los organismos, algunos están interesados principalmente en las actividades de las moléculas que componen las células nerviosas y otros dedican su atención a los fenómenos intermedios. Los campos de la biología

molecular y la neurobiología, no suelen retroceder al nivel cuántico para hablar de las actividades de, por ejemplo, los enlaces químicos. Rara vez los biólogos son impulsados por anomalías o por cualquier otra razón para llegar a niveles tan bajos, aunque algún problema podría requerirlo. Los niveles por debajo de las moléculas y los enlaces químicos no son fundamentales para los campos de la biología molecular, pero sí para la nanobiología³⁹.

Pero recuerde, lo que se considera el nivel de fondo puede cambiar. En biología molecular y neurobiología molecular, las jerarquías de mecanismos tocan fondo en las descripciones de las actividades de macromoléculas, moléculas más pequeñas y de iones. Estos se reconocen comúnmente como entidades de fondo; creemos que hemos identificado los tipos más importantes de actividades de fondo. Estas actividades de fondo en biología pueden clasificarse en cuatro tipos: (i) geométrico-mecánico; (ii) electroquímica; (iii) energética; (iv) electromagnética. (i) Las actividades geométrico-mecánicas son aquellas familiares de la filosofía mecánica del siglo XVII. Incluyen encajar, girar, abrir, chocar, doblar y empujar. La rotación de la hélice alfa en el canal de sodio y el ajuste geométrico de un neurotransmisor y un receptor postsináptico son ejemplos de actividades geométrico-mecánicas. (ii) Atraer, repeler, unir y romper son tipos de actividad electroquímica. Los enlaces químicos, como la formación de enlaces covalentes fuertes entre los aminoácidos en las proteínas, es un ejemplo más específico. El acoplamiento de la cerradura y la llave de una enzima y su sustrato implica formas geométricas y tensiones mecánicas y atracciones químicas. Como veremos, el desarrollo histórico del mecanismo de síntesis de proteínas requirió encontrar una actividad para ordenar linealmente los constituyentes de la proteína, sus aminoácidos; una idea temprana que utilizaba principalmente actividades geométrico-mecánicas fue reemplazada por otra que involucraba, principalmente, las débiles actividades electroquímicas de los enlaces de hidrógeno. (iii) Las actividades energéticas tienen la termodinámica como fuente. Un tipo de actividad energética implica la simple difusión de una sustancia, como, por ejemplo, cuando las concentraciones en diferentes lados de una membrana conducen al movimiento de sustancias a través de la membrana. (iv) Las actividades electromagnéticas se utilizan ocasionalmente para tocar fondo a los mecanismos en estas ciencias. La conducción de impulsos eléctricos por las células nerviosas y los mecanismos de navegación de ciertas especies marinas son ejemplos.

La comprensión actual de las ciencias naturales básicas (incluidas la física y la química) se centra en la energía y la masa como aspectos principales de cualquier materia (material) del universo. La correlación entre la energía y la masa como dos aspectos principales de la materia ha sido descrita claramente a través de la ecuación de la relatividad de la energía y la masa de Albert Einstein. Sin embargo, aún, la posibilidad de almacenamiento de información en moléculas (incrustación de código en moléculas) queda por describir sistemáticamente como la tercera dimensión de la materia. Las biomoléculas como los ácidos nucleicos son modelos prometedores para la descripción de la posibilidad de almacenamiento de información en la materia (capacidad codificante de la materia). Aquí, primero discutimos los fundamentos de la química y la física en los sistemas vivos y no vivientes. Luego, al referirnos a la posibilidad de almacenamiento de información (código) en moléculas, introducimos el código como la tercera dimensión de la materia (por ejemplo, códigos incrustados en biomoléculas). Al considerar la capacidad de almacenamiento de información de las biomoléculas, describimos esta propiedad funcional tridimensional de las biomoléculas (y cualquier otra molécula programable) mediante la relatividad del código, la energía y la masa. De hecho, al aprender de las propiedades autónomas de los sistemas biológicos y recapitular la capacidad de codificación de las biomoléculas, definimos una nueva traducción de la ecuación de la relatividad como la relatividad del código, la energía y la masa. La relatividad del código, la energía y la masa como concepto emergente en las ciencias naturales revela la lógica molecular detrás de varias características de la naturaleza, incluidas las propiedades autónomas de los sistemas biológicos. Más allá de las ciencias biológicas, la aplicación de los principios básicos de la relatividad del código, la energía y la masa puede abrir nuevas vías hacia la generación de sistemas autónomos sintéticos con una amplia gama de aplicaciones en diferentes áreas de la ciencia, incluidas la informática, la energía, la física y las ciencias de nanomateriales.

0.9 Los sistemas biológicos son altamente autónomos

Los organismos vivos son sistemas autónomos altamente avanzados. Por definición, una característica clave de la autonomía tanto en los sistemas naturales como en los artificiales está relacionada con la capacidad de funcionamiento independiente de los procesos⁴⁰. Basado en la teoría del cataclismo cósmico, que es la teoría más popular sobre el origen del universo, hace unos 14 mil millones de años, el universo se formó como una erupción cataclísmica masiva de partículas subatómicas calientes y ricas en

energía. La teoría sostiene que después de la erupción, el universo se expandió. A medida que pasaba el tiempo y la materia se enfriaba, comenzaron a formarse tipos más diversos de átomos que eventualmente se condensaron y produjeron estrellas y galaxias de nuestro universo actual, incluida la Tierra y todos sus elementos químicos⁴¹. Aunque el origen de la vida no se puede detectar con precisión, según la evidencia y los residuos de organismos similares a bacterias encontrados en la investigación geológica, se estima que la vida se inició en la tierra alrededor de 3.5 mil millones de años atrás. Incluso antes, hace casi 4 mil millones de años, existían componentes muy simples de la vida, como las moléculas funcionales primitivas. Con el tiempo, estas moléculas funcionales se conectaron entre sí y formaron formas de vida primitivas⁴². La biología es la ciencia natural que involucra el estudio de las propiedades físicas, químicas, funcionales y de desarrollo de la vida. Los organismos vivos (también llamados sistemas biológicos) se consideran sistemas autónomos de gran complejidad⁴³. Como característica notable de la naturaleza, los sistemas biológicos poseen varios aspectos de autonomía⁴⁴:

1. Los sistemas biológicos se autoensamblan. El **autoensamblaje** es un proceso en el que un sistema desordenado con muchos componentes se convierte en una estructura ordenada y estable (en equilibrio con configuración mínima de energía) sin dirección externa. Por lo general, involucra solo los enlaces más débiles (por ejemplo, enlaces de hidrógeno e interacciones hidrofóbicas), lo que hace que la estructura del sistema sea más flexible. Un ejemplo de autoensamblaje en sistemas no vivos es la formación de cristales. El plegamiento de cadenas polipeptídicas en proteínas y ácidos nucleicos en sus formas funcionales son ejemplos de estructuras biológicas autoensambladas.

2. Los sistemas biológicos se autoorganizan. La autoorganización significa que un sistema tiene la capacidad de ensamblar sus propias subunidades a través de múltiples capas y escalas de complejidad. El desarrollo embrionario es un ejemplo típico de procesos de autoorganización en sistemas biológicos⁴⁵. Durante las primeras etapas del desarrollo embrionario, las células madre son capaces de ensamblar y formar estructuras muy específicas que crecen en tamaño y complejidad con el tiempo⁴⁶. Si bien tanto los procesos de autoensamblaje como de autoorganización se refieren a la formación espontánea de una estructura ordenada, se pueden distinguir entre sí en diferentes aspectos. Mientras que el autoensamblaje llega a una configuración de equilibrio, la autoorganización aleja al sistema del equilibrio que requiere una infusión

de energía para mantenerlo. Otra gran diferencia es la interacción. Si bien la interacción en el autoensamblaje es más elemental, lo que se remonta a la fuerza electromagnética, la interacción en la autoorganización es más compleja. El tamaño de la estructura resultante del autoensamblaje es principalmente mesoscópico, pero puede ser considerablemente mayor para la autoorganización⁴⁷. En resumen, el autoensamblaje se puede observar tanto en sistemas vivos como no vivos, mientras que la autoorganización es una propiedad especial de los sistemas vivos.

3. Los sistemas biológicos se auto-repican. La **autorreplicación** significa que un sistema puede producir una copia precisa de sí mismo. Todos los sistemas vivos, independientemente de su nivel de desarrollo, desde las bacterias más primitivas hasta los organismos eucariotas más desarrollados, pueden hacer una copia fiel de sí mismos.

4. Los sistemas biológicos son de **autorreparación**, significa que un sistema tiene la capacidad de regenerar directamente su propia parte dañada por sí mismo, independientemente de cualquier mediador u operador externo.

5. Los sistemas biológicos se alimentan por sí mismos. Autoalimentación significa que un sistema tiene la capacidad de producir y absorber directamente sus propias necesidades energéticas de su entorno, sin la necesidad de ningún mediador u operador externo⁴⁸.

6. Los sistemas biológicos indican el comportamiento de la maquinaria en múltiples niveles de complejidad. Además del autoensamblaje y la autoorganización que conducen a la formación de estructuras altamente específicas, se pueden formar miles de maquinarias durante el progreso de un proceso de autoorganización en sistemas biológicos. A diferencia de cualquier otro sistema creado por el hombre, los sistemas biológicos están hechos de millones de maquinarias (módulos) a diferentes escalas, que van desde nano hasta micromecanismos y meso maquinarias, que se integran entre sí para formar **módulos funcionales** de un sistema biológico. Por ejemplo, ADN polimerasa hasta orgánulos celulares como el ribosoma como parte de la maquinaria de síntesis de proteínas, tejidos y órganos, por ejemplo, el riñón es una maquinaria muy avanzada de filtración.

7. Los sistemas biológicos son autocontrolados en su interacción con los factores ambientales. Autocontrol significa que un sistema tiene la capacidad de codificar, controlar y ejecutar su propio procesador⁴⁹. Los sistemas biológicos aplican sensores y receptores moleculares muy especializados para la interacción dinámica con sus factores ambientales. Un ejemplo de comportamiento de autocontrol en un sistema biológico de **entrelazamiento cuántico** que incluye la capacidad de auto-navegación de las aves durante su vuelo migratorio. La existencia de proteínas magnetorreceptoras en sus ojos puede ser responsable de la capacidad de detectar el campo magnético de la tierra y usarlo para navegar en direcciones en destinos cortos y largos⁵⁰; en la plasticidad de plantas para jugar con el ángulo de la radicación solar⁵¹.

8. Los sistemas biológicos tienen la capacidad de detectar y responder a factores ambientales. Las capacidades de detectar y responder a los factores ambientales (o señales) conducen a la conducta cognitiva en un sistema y en la forma avanzada permiten adquisición, percepción, comprensión y recuerdo de información.

Dado que los sistemas biológicos están altamente conectados con los elementos inanimados y los sistemas no vivos de la naturaleza, quedan por responder varias preguntas desafiantes.

- ¿Cuáles son los mecanismos reguladores detrás de las propiedades autónomas de los sistemas biológicos que los distinguen de los elementos inanimados y los sistemas no vivos en la naturaleza?
- ¿Somos capaces de explicar todas las propiedades autónomas de los sistemas biológicos aplicando las leyes naturales definidas actualmente en física y química?

1.3 Importancia y aplicaciones fundamentales de los mecanismos reguladores de la autonomía en los sistemas biológicos

La comprensión profunda de los mecanismos reguladores detrás de las propiedades autónomas de los sistemas biológicos es importante en varios aspectos, entre ellos:

1. Considerar el cuerpo humano (incluidos todos los órganos internos y el cerebro) como un sistema autónomo altamente avanzado, toda enfermedad es el resultado de un desorden en los mecanismos y programas reguladores que están ejecutando este sistema autónomo.

2. La comprensión profunda de los programas moleculares detrás de la maquinaria del cuerpo humano es fundamental para la invención de nuevos métodos terapéuticos, especialmente enfoques para enfermedades actualmente incurables (por ejemplo, cáncer, trastornos genéticos, Alzheimer, Parkinson y esclerosis múltiple).

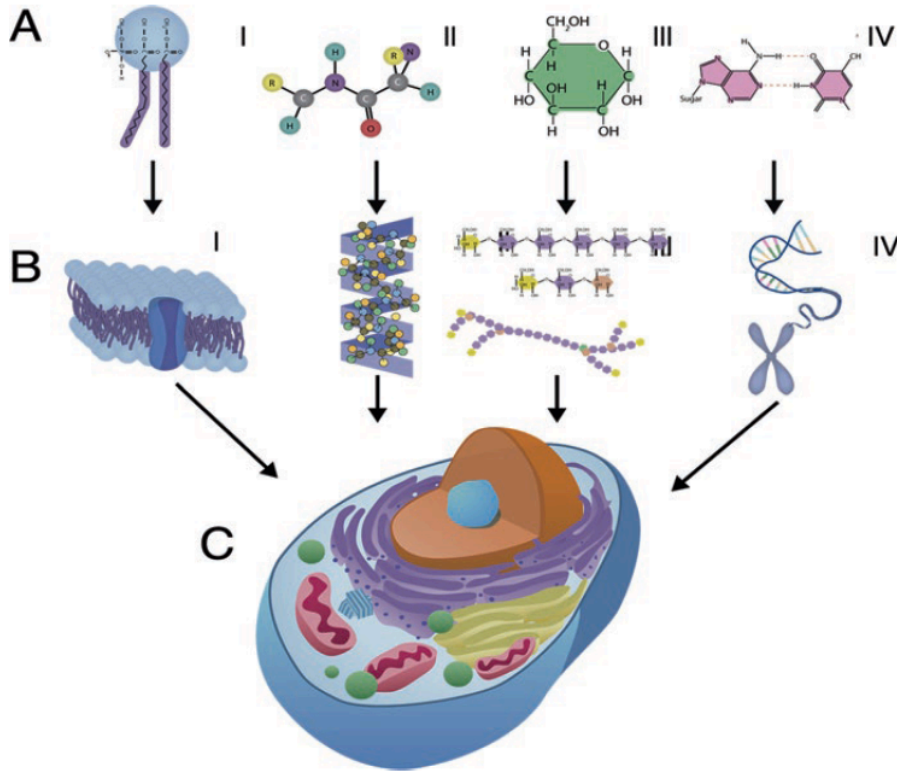
3. La creciente demanda de creación de sistemas de inteligencia artificial (IA) capaces de conectarse directamente con el cerebro humano requiere una comprensión profunda de los mecanismos moleculares del procesamiento autónomo de datos en el sistema neuronal humano. Esto es fundamental para la generación de algoritmos inteligentes innovadores que podrían usarse para una amplia gama de aplicaciones. Por ejemplo, la simulación de mecanismos de procesamiento de datos autónomos en el cerebro humano se puede aplicar para generar sistemas de IA portátiles (por ejemplo, interfaces cerebro-máquina humano) para diferentes propósitos terapéuticos o para mejorar el rendimiento del cerebro humano.

4. Dado que los sistemas biológicos son sistemas autónomos naturales muy avanzados, la comprensión y la ingeniería inversa de su sistema operativo se pueden aplicar para diseñar sistemas inteligentes sintéticos en una amplia gama de industrias, como dispositivos electrónicos inteligentes, materiales inteligentes y dispositivos inteligentes biocompatibles y portátiles.

0.9.1 Fundamento químico de la vida

Al principio, puede resultar difícil aceptar la idea de que cada criatura viviente es simplemente un sistema químico⁵². Hasta el siglo XIX, se creía que los animales contenían una fuerza vital - "un animus" - que era responsable de sus propiedades distintivas. Sin embargo, científicamente, los sistemas biológicos solo están compuestos por elementos químicos naturales. Las células, como subunidad principal de los sistemas biológicos, están formadas por la organización jerárquica de moléculas. La célula ilustra la jerarquía estructural en la organización. La química de los sistemas biológicos (desde las formas más primitivas hasta los organismos altamente avanzados) se componen principalmente de cuatro elementos naturales: **carbono, hidrógeno, oxígeno y nitrógeno**, que normalmente constituyen hasta el 96% de la masa seca de las células vivas. Aparte de los cuatro elementos antes mencionados, los siguientes

elementos clave de los sistemas biológicos son **fósforo, azufre, sodio, cloruro, potasio, calcio y magnesio**, que constituyen hasta el 3.5% de la masa seca de la célula⁵³.



Además, los oligoelementos están presentes en niveles muy bajos (menos del 0.5% de la masa seca) en un organismo. Estos oligoelementos son esenciales para la vida y los sistemas vivos no pueden sobrevivir sin ellos. Los oligoelementos incluyen hierro, yodo, manganeso, molibdeno, selenio, silicio, estaño, vanadio, boro, cromo, cobalto, zinc, cobre y flúor⁵⁴. Los oligoelementos son esenciales para el sustento de los organismos vivos debido a su papel fundamental en la función de proteínas específicas. Por ejemplo, el yodo es importante para producir diferentes formas de hormona tiroidea e implica la regulación del crecimiento y los niveles de energía en los mamíferos. Muchos de los oligoelementos son necesarios para la función adecuada de las enzimas. El hierro, que se encuentra en los glóbulos rojos, ayuda a transportar oxígeno a diferentes tejidos. Entre los cuatro elementos principales de las células vivas, el carbono constituye más de la mitad del peso seco de las células y desempeña el papel más importante en los organismos vivos. El papel clave del carbono en los sistemas biológicos está relacionado con su alta flexibilidad para formar enlaces químicos

consigo mismo y con otras moléculas. El carbono puede formar enlaces sencillos con átomos de hidrógeno y enlaces sencillos y dobles con átomos de oxígeno y nitrógeno. Una de las propiedades más importantes en biología es la capacidad de los átomos de carbono para formar enlaces sencillos carbono-carbono muy estables. Cada átomo de carbono puede formar enlaces sencillos con hasta otros cuatro átomos de carbono. Dos átomos de carbono también pueden formar enlaces dobles o triples. Por tanto, los átomos de carbono unidos covalentemente pueden formar una amplia gama de estructuras moleculares que incluyen cadenas lineales, cadenas ramificadas y estructuras cíclicas. Los grupos de átomos reactivos como el oxígeno, el nitrógeno y el fósforo, también llamados grupos funcionales, se agregan al esqueleto de carbono principal para formar diferentes tipos de biomoléculas. La existencia de grupos funcionales en la estructura química de las biomoléculas determina la reactividad, solubilidad y otras propiedades físicas de una biomolécula y define su papel biológico en un organismo vivo⁵⁵. Los alcoholes son biomoléculas que contienen al menos un grupo hidroxilo. Con una fuerte electronegatividad de los átomos de oxígeno, los grupos hidroxilo proporcionan polaridad para las biomoléculas. De manera similar, los grupos carbonilo en aldehídos y cetonas generalmente aumentan la polaridad y la reactividad de las moléculas orgánicas. Los grupos carboxilo, que contienen dos átomos de oxígeno en su estructura, son grupos funcionales altamente polares y reactivos. Las aminas contienen grupos amino. Al igual que el amoníaco libre (NH_3), las aminas son débilmente básicas y pueden unirse a iones de hidrógeno electropositivos en una solución. Por lo tanto, las aminas pueden formar enlaces de hidrógeno con agua y otras moléculas polares en una solución acuosa. En el extremo opuesto del espectro, los grupos fosfato son muy ácidos y se encuentran principalmente en un estado ionizado en soluciones acuosas liberando iones hidrógeno. Debido a los cuatro átomos de oxígeno electronegativos, los grupos fosfato son extremadamente reactivos. La transferencia de grupos fosfato entre moléculas puede aportar energía a las reacciones químicas. Por ejemplo, el trifosfato de adenosina (ATP), que contiene tres grupos fosfato en su estructura, funciona como el principal portador de energía en los sistemas biológicos⁵⁶.

Muchas biomoléculas son polifuncionales y contienen dos o más grupos funcionales diferentes en sus estructuras químicas. Esta complejidad estructural de moléculas polifuncionales proporciona una flexibilidad especial para que estas moléculas cambien entre diferentes roles según la demanda del sistema.

0.9.2 Fundamento físico de la vida

Las células vivas deben realizar un trabajo para mantener la organización de sus moléculas, reproducirse y, en general, seguir viviendo. Las reacciones sintéticas que ocurren dentro de la célula, como los procesos sintéticos en cualquier fábrica, requieren el aporte de energía. También se consume energía para las operaciones de movimiento y maquinaria de las células vivas. Procesar y almacenar información también requiere energía, sin la cual las estructuras ricas en información se desordenan. Para las reacciones químicas que ocurren en una solución, podemos definir un sistema como todas las reacciones y productos constituyentes: el solvente que los contiene y la atmósfera inmediata dentro de una región definida del espacio. Por lo tanto, un sistema biológico se considera un sistema abierto: intercambia energía y materia (masa) con su entorno. Los sistemas biológicos obtienen energía de su entorno de dos formas:

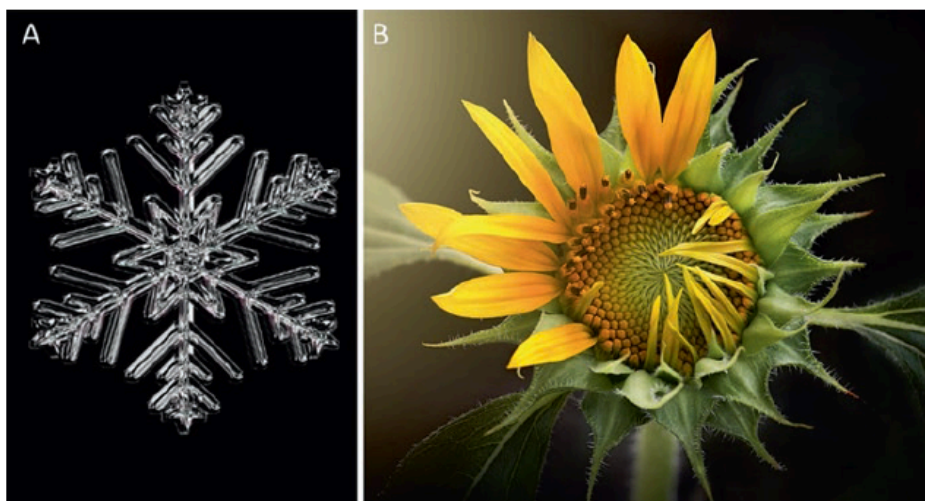
1. Captan energía química (como la glucosa) de su entorno y extraen energía por oxidación.
2. Absorben energía de la luz solar.

La primera ley de la termodinámica describe el principio de conservación de la energía: en cualquier cambio químico, la cantidad total de energía en el universo permanece constante, aunque la forma de la energía puede cambiar. Además, según la conocida ley física de la relatividad general, explicada por Albert Einstein, la energía y la masa son relativas. Cada materia (masa), incluso en forma estática, contiene un nivel especial de energía que se almacena en sus moléculas. La relatividad de la energía y la masa se explica por la ecuación de relatividad $E = mc^2$, donde “m” es la masa y “c” es la velocidad de la luz (aproximadamente⁵⁷ $3 \times 10^8 m/s$). Además, a diferencia de la física clásica donde la luz se consideraba simplemente como energía en forma de onda, Einstein afirmó que la luz es una partícula que contiene energía correspondiente a su longitud de onda. Einstein logró explicar el efecto fotoeléctrico, que no se puede explicar si solo se considera la luz como una onda. Entonces, la luz de esta manera pasó a llamarse un “fotón (cuantos de luz)” ya que tiene las propiedades no solo de una onda sino también de una partícula. La ecuación de relatividad fue inicialmente escrita

en muchas notaciones diferentes. Se desarrollaron más interpretaciones y justificaciones de la fórmula para diferentes aplicaciones⁵⁸. Inicialmente, la eq. $E = mc^2$ no se escribió como una fórmula, pero Einstein la describió en una oración, de la siguiente manera, “si un cuerpo emite la energía L (en forma de radiación), su masa disminuye, se puede calcular como⁵⁹ L/V^2 . Posteriormente, describió la equivalencia en la forma⁶⁰ $M = \mu + E0/C^2$. Después de la Segunda Guerra Mundial, Einstein volvió a la ecuación de la relatividad una vez más y esta vez escribió $E = mc^2$ en el título de un artículo⁶¹. Intentó describir la ecuación de la relatividad en una forma más general.

Posteriormente, también describió la equivalencia de masa y energía, claramente como sigue: la energía E de los sistemas físicos numéricamente igual a la producción de su masa (m) y la velocidad de la luz C al cuadrado. La equivalencia masa-energía establece que un objeto tiene una cierta cantidad de energía, incluso cuando está estacionario. Según la mecánica newtoniana, un cuerpo estacionario (inmóvil) no tiene energía cinética y puede tener o no otras cantidades de energía almacenada internamente. Mientras que, según la relatividad de la energía y la masa, cualquier objeto tiene una energía equivalente (E) que se puede calcular como la masa (m) multiplicada por la velocidad de la luz ($c =$ aproximadamente $3 \times 10^8 m/s$) al cuadrado⁶². De manera similar, todavía hay varios fenómenos desconcertantes en la naturaleza observados en las propiedades químicas y físicas de los sistemas biológicos que no se pueden explicar si uno solo considera la energía y la masa como los únicos aspectos de la materia. Los sistemas biológicos obedecen las leyes naturales definidas actualmente en física, incluida la termodinámica, la relatividad de la energía y la masa. Sin embargo, no es suficiente para explicar todas las propiedades de los sistemas biológicos. Por ejemplo, las propiedades autónomas de los sistemas biológicos (incluida su capacidad de autoorganización y autorreplicación) no se pueden explicar simplemente si consideramos solo la energía y la masa. La autoorganización es una propiedad notable en la química de los sistemas biológicos que los distingue de los sistemas no vivos de la naturaleza. A diferencia del autoensamblaje que se puede observar tanto en sistemas vivos como no vivos (por ejemplo, formación de cristales de hielo o plegamiento espontáneo de proteínas), la autoorganización solo se puede observar en sistemas vivos.

El autoensamblaje es un proceso espontáneo y, como reacción química exergónica, no requiere la absorción de energía del medio ambiente. Por tanto, el autoensamblaje puede explicarse por la termodinámica y por la estabilidad de la estructura molecular. Sin embargo, la formación y el mantenimiento de un sistema autoorganizado dependen de la entrada continua de energía en el sistema. Por lo tanto, al considerar el balance energético del sistema, la autoorganización requiere la absorción de energía del medio ambiente. Como otra propiedad natural notable, los sistemas biológicos se autoalimentan y pueden gestionar sus necesidades energéticas. Otra propiedad notable de un sistema autoorganizado es la capacidad de romper la simetría estructural que también se puede definir como formación de polaridad estructural⁶³. La formación de polaridad implica la generación autónoma y el ensamblaje de estructuras geométricas altamente específicas, que crecen a través de diferentes capas de complejidad desde nanoestructuras hasta micro y macroestructuras.



Por lo tanto, la formación de polaridad en un sistema autoorganizado requiere programas internos para orientar y reforzar el progreso hacia la formación de una estructura geométrica deseada. En otras palabras, la autoorganización solo ocurre a través de un proceso programado internamente dedicado. Aunque tanto los procesos de autoensamblaje como de autoorganización se refieren a la formación espontánea de una estructura ordenada, se diferencian entre sí en varios aspectos. En primer lugar, la autoorganización es un **proceso no lineal**. Durante la autoorganización, se pueden formar diferentes tipos de estructuras en múltiples niveles de complejidad. Mientras que la interacción molecular en el autoensamblaje es más elemental y se remonta a la

fuerza electromagnética. El tamaño de la estructura resultante del autoensamblaje es principalmente mesoscópico, pero puede ser considerablemente mayor para la autoorganización. Teniendo en cuenta la base química de los sistemas biológicos, se requieren programas reguladores detrás de la autoorganización y deben integrarse directamente dentro de la química del sistema. Esto implica las interacciones electroquímicas entre moléculas, que también se pueden definir como codificación molecular endógena. Sin embargo, la capacidad de codificación de las moléculas, así como la programabilidad de un sistema químico a través de las interacciones moleculares jerárquicas, todavía no se ha explicado claramente por las leyes de la termodinámica ni la relatividad de la energía y la masa.

La explicación científica de la autoorganización en un sistema biológico requiere la consideración de los programas endógenos (códigos), que están incrustados en la estructura fisicoquímica de las moléculas, como un factor adicional más allá de la energía y masa del sistema. Los códigos moleculares endógenos, en un sistema autoorganizado, guían el progreso del proceso a través de múltiples capas de complejidad. La posibilidad de codificación molecular endógena se recapituló a partir de las propiedades de codificación de biomoléculas. Por lo tanto, aquí describimos esta propiedad de codificación molecular de las biomoléculas como la tercera dimensión de las leyes de la naturaleza en química y física.

0.9.3 Código: la tercera dimensión de la ley de la naturaleza

La comprensión actual de las ciencias naturales básicas (incluidas la física y la química) se centra en la energía y la masa como aspectos principales de cualquier materia (material) del universo. La correlación entre la energía y la masa como dos aspectos principales de la materia ha sido descrita claramente a través de la ecuación de la relatividad de la energía y la masa de Albert Einstein. Sin embargo, aún, la posibilidad de almacenamiento de información en moléculas (incrustación de código en moléculas) queda por describir sistemáticamente como la tercera dimensión de la materia. Las biomoléculas como los ácidos nucleicos son modelos prometedores para la descripción de la posibilidad de almacenamiento de información en la materia (capacidad codificante de la materia). Aquí, describimos la posibilidad de incrustar información (código) en la estructura fisicoquímica de las moléculas, como la tercera dimensión de la ley de la naturaleza (más allá de la energía y la masa). Los sistemas biológicos se

definen como sistemas programados con precisión mediante el almacenamiento y la operación de información por parte de sus moléculas. Los ácidos nucleicos son moléculas bien conocidas que funcionan como sustancias químicas codificadas altamente conservadas en la naturaleza. El ácido desoxirribonucleico (ADN) es el material hereditario en los seres humanos y en casi todos los demás organismos⁶⁴. La información en el ADN se almacena como un código compuesto por cuatro bases químicas: adenina (A), guanina (G), citosina (C) y timina (T). El orden y las secuencias de estas bases determinan la información disponible para construir y mantener un organismo. Además de la función de codificación biológica del ADN, también se ha utilizado como matriz para el almacenamiento y procesamiento de información no biológica en forma sintética⁶⁵. Los resultados de esta investigación se extendieron a la capacidad de codificación de moléculas de ADN desde la biología hasta las matemáticas y la informática. Por ejemplo, las moléculas de ADN sintético se han utilizado como una matriz de cálculo para resolver un grupo de problemas matemáticos complejos llamados problemas de tiempo polinomiales no deterministas (NP). Los problemas NP son una clase de problemas matemáticos con complejidad exponencial sin una solución eficiente actual. Dada la limitación de las actuales computadoras basadas en silicio en funcionamiento secuencial, es casi improbable que los problemas de NP se resuelvan en un tiempo lógico⁶⁶. Los resultados de los estudios actuales sobre computación con moléculas de ADN indican que el ADN tiene muchas ventajas sobre los métodos computacionales convencionales que utilizan semiconductores de estado sólido.

Las moléculas de ADN pueden ejecutar miles de millones de operaciones simultáneamente y en paralelo entre sí. Esto contrasta con las computadoras digitales eléctricas donde las operaciones individuales son muy rápidas; sin embargo, las operaciones se ejecutan secuencialmente. El paralelismo masivo de la computación del ADN proviene de la gran cantidad de moléculas que interactúan químicamente en un pequeño volumen. Las moléculas de ADN también proporcionan una gran capacidad de almacenamiento en un espacio muy pequeño porque codifican información a nivel molecular. Basado en el paralelismo masivo, la computación de ADN proporciona nuevas soluciones para los problemas de NP. El ADN es un ejemplo de materiales programables naturales. Sin embargo, además del ADN, se pueden aplicar varios otros materiales naturales y sintéticos para la codificación y el procesamiento de datos en nanocomputadoras químicas de próxima generación. Aquí, al aprender de la naturaleza y considerar la posibilidad de almacenamiento y operación de información

por moléculas codificadas, así como la interacción programable entre moléculas entre sí y sus factores ambientales, discutimos tres nuevos conceptos emergentes en ciencias naturales, de la siguiente manera:

1. Química programable
2. Química algorítmica
3. Química cognitiva

La **química programable** es una forma especial de química en la que la información está incrustada en enlaces químicos de moléculas o interacciones electroquímicas entre diferentes moléculas. En otras palabras, las moléculas se pueden aplicar como una matriz para el almacenamiento de información en un sistema químico programable. La **química algorítmica** es una forma especial de química en la que la información puede procesarse mediante una secuencia de operaciones moleculares. Esto significa que, en un sistema de química algorítmica, un grupo de moléculas están interconectadas y pueden funcionar como un algoritmo. De hecho, el sistema puede funcionar como un algoritmo mediante una secuencia de interacciones electroquímicas entre moléculas. Por tanto, más allá de la capacidad de codificación de las moléculas, en un sistema de química algorítmica, la información puede ser procesada por un conjunto de operadores moleculares programados internamente. La **química cognitiva** es una forma especial de química en la que el sistema es capaz de reconocer y responder a las señales ambientales dinámicas e incluso adaptarse a sus condiciones ambientales. El sistema está altamente conectado con sus factores ambientales a través de sensores moleculares y puertas lógicas. El sistema está formado por un grupo de moléculas codificadas con precisión (que a través de sus interacciones electroquímicas) están altamente conectadas entre sí, así como con sus factores ambientales. De hecho, un sistema de química cognitiva tiene la capacidad de cognición endógena. En los siguientes capítulos, proporcionaremos una descripción detallada de cómo se puede aplicar la química cognitiva en formas sintéticas como una herramienta emergente para los sistemas de computación basados en sustancias químicas de próxima generación. Además, discutiremos cómo se puede aplicar la aplicación de los principios básicos de la relatividad del código, la energía y la masa para replicar las propiedades notables de los sistemas biológicos, incluida la autoorganización, la autorregulación, la toma de decisiones, etc.

0.9.4 Relatividad de código, energía y masa

Aquí, al considerar la posibilidad de almacenamiento de información en moléculas, describimos una nueva traducción de la ecuación de relatividad, como la relatividad de código, energía y masa. Una nueva traducción de la ecuación de la relatividad explica que, en la naturaleza, la materia no se limita a la energía y la masa, y la transformación no se limita solo a la energía y la masa, sino que también incluye el código como factor adicional. La relatividad del código, la energía y la masa como una nueva plataforma en la ciencia revela la lógica molecular detrás de varias características desconocidas de la naturaleza (como la maquinaria molecular natural y la autoorganización) y brinda soluciones para varios desafíos en diferentes áreas de la tecnología y las ciencias de la vida. Como ejemplo, primero explicamos la existencia triple de la relatividad en moléculas de ácidos nucleicos (por ejemplo, inicialmente en moléculas de trifosfato de adenosina). Sin embargo, se puede explicar para todas las moléculas naturales o sintéticas con capacidades multifuncionales similares de codificación, energía y transformación de masa a través de sus estructuras moleculares (por ejemplo, péptidos sintéticos, grafeno y aptámeros). Dada la naturaleza multifuncional de los ácidos nucleicos, estas moléculas también pueden actuar como portadores a corto plazo de energía química en sistemas biológicos. Por encima de todos los demás, el trifosfato de adenosina ribonucleico (ATP) transfiere energía en cientos de reacciones celulares diferentes. El ATP se puede formar mediante el reciclaje de la energía liberada por la degradación oxidativa de otras biomoléculas. De hecho, los ácidos nucleicos son moléculas que la naturaleza ha seleccionado como portadores de energía. Por ejemplo, el ATP se puede sintetizar mediante las reacciones de fosforilación oxidativa de la fotosíntesis, que implica la conversión de la energía solar en sus enlaces químicos. La fotosíntesis es un representante de un proceso natural que ilustra claramente la existencia y precisión de la relatividad del código, la energía y la masa en la naturaleza mediante la conversión de la energía solar en energía química.

Tres fosfatos de la molécula de ATP están unidos en serie por dos enlaces fosfoanhidrato, cuya ruptura libera grandes cantidades de energía útil. En particular, el grupo terminal de fosfato se escinde con frecuencia por hidrólisis, a menudo transfiriendo un fosfato a otras moléculas y liberando energía que impulsa los requisitos de energía para las reacciones biosintéticas. Otros derivados de nucleótidos (por ejemplo, GTP, NADH) también funcionan como portadores de energía para otros

grupos químicos. La fotosíntesis es un proceso natural que ilustra claramente cómo un sistema químico programado puede generar una maquinaria molecular para transformar la energía solar en una masa de biomoléculas. Esta energía también se puede aplicar a la generación de nuevos materiales codificados y la autoorganización de nuevos sistemas cognitivos. La integración de la capacidad de codificación y las propiedades termodinámicas (transformación de energía y masa) de las biomoléculas conduce a la alta flexibilidad y propiedades multifuncionales de estas moléculas.

Considerando los fundamentos físicos y químicos de la vida, se puede concluir que los organismos vivos son sistemas fisicoquímicos y no están exentos de las leyes de la naturaleza en física y química. Científicamente sabemos que los sistemas biológicos están compuestos por elementos normales de la naturaleza. Sin embargo, los mecanismos reguladores que por primera vez provocaron la conversión de elementos de la aleatoriedad en un sistema altamente organizado (determinado como vida) siguen siendo un misterio. “La ciencia no puede resolver el misterio ilimitado de la naturaleza. Y es que, en última instancia, nosotros mismos somos parte de la naturaleza y, por tanto, parte del misterio que estamos tratando de resolver ” (Max Planck). Los mecanismos reguladores precisos detrás del "origen de la vida" hace cinco mil millones de años, cuando ocurrió la primera reacción autogeneradora de la vida, seguían siendo un misterio de la naturaleza (debido a nuestras limitaciones en la experiencia del tiempo). Sin embargo, el lado positivo es que podemos aprender de las reglas de vida actualmente existentes que ya son accesibles a la experiencia humana. A principios del siglo XX, justo después del establecimiento de la física moderna y el efecto revolucionario de la física cuántica en diferentes tecnologías, Erwin Schrodinger se refirió al surgimiento de una revisión en la definición de vida. Además, intentó describir la revisión en la clasificación de las ciencias naturales llenando los vacíos entre la vida y otras ramas de las ciencias. En su conocido libro sobre la vida, titulado “¿Qué es la vida?”. Se refirió a las partes faltantes en la clasificación de las diferentes ramas de las ciencias naturales que separan los sistemas vivos de otros sistemas fisicoquímicos creados por el hombre. Describió que esta brecha está relacionada con nuestra comprensión limitada de la estructura y construcción de los sistemas vivos. Dicho de otra manera, la construcción de sistemas vivos es diferente de cualquier cosa que hayamos probado hasta ahora en los laboratorios físicos hechos por humanos. “Lo que deseamos aclarar es que de todo lo que hemos aprendido sobre la estructura de la materia viva, debemos estar preparados para encontrarla funcionando de una manera

que no pueda reducirse a las leyes ordinarias de la física. Eso no porque haya una nueva “fuerza” o lo que sea, que dirija el comportamiento de los átomos individuales dentro de un organismo vivo, sino porque la construcción es diferente de cualquier cosa que hayamos probado hasta ahora en el laboratorio físico⁶⁷. A principios del siglo XX, junto con el gran progreso en física y química y más tarde en matemáticas (ciencia de datos), ocurrió un movimiento revolucionario en las clasificaciones convencionales de ciencias naturales. Diferentes ramas de las ciencias naturales, incluidas la física, la química y las matemáticas, se acercaron unas a otras y, posteriormente, se establecieron varios campos nuevos en las ciencias naturales. Por ejemplo, la física moderna (cuántica) se estableció mediante una nueva interpretación de la física y la química integradas. Más tarde, la informática se estableció mediante la integración de la física electrónica y la ciencia de datos (matemáticas). Durante la Segunda Guerra Mundial, Alan Turing desarrolló la primera computadora digital universal (también llamada máquina de Turing). Trabajó en álgebra y teoría de números, así como en una máquina de cifrado basada en relés electromagnéticos para multiplicar números binarios.

A principios del siglo XX, Max Planck estableció los principios básicos de la física moderna. Más tarde, la física moderna se volvió muy práctica y se aplicó en diferentes tecnologías mediante el trabajo de investigación de Albert Einstein y Erwin Schrodinger sobre la mecánica cuántica. A las 1900, Max Planck intentó explicar la radiación del cuerpo negro de los objetivos calientes. Luego, el espectro de un cuerpo negro se cuantificó asumiendo que la energía en una onda electromagnética, según Max Planck eq. ($E = nhf$). En 1905, Albert Einstein explicó el concepto de fotón como una bolsa de radiación electromagnética que se comporta como energía (onda) y partícula (masa) al mismo tiempo. Además, describió la ecuación para la relatividad de la energía y la masa como $E = mc^2$. Más tarde, en 1927, Erwin Schrodinger intentó definir una ecuación de onda que pudiera describir con precisión la energía en una partícula de átomo. Schrodinger logró presentar una ecuación que podía predecir con precisión las energías en los átomos de hidrógeno, lo que llevó al establecimiento de los principios básicos de la física cuántica. La ecuación de Schrodinger probablemente fue la ecuación más importante del siglo XX, debido a su enorme efecto sobre el progreso tecnológico. La mayor importancia del trabajo de Schrodinger en física cuántica proviene de su amplia cobertura tanto de la física como de la química. El físico Paul

Dirac afirmó: la ecuación de Schrodinger explica gran parte de la física y toda la química. Además, a mediados del siglo XX, Schrodinger intentó revisar la definición clásica de ciencias de la vida mediante la aplicación de principios básicos de física y química en sistemas biológicos. El manuscrito completo de su serie de conferencias sobre la base fisicoquímica de la vida se publicó más tarde como un libro titulado "¿Qué es la vida?". Schrodinger intentó proporcionar varias pruebas para demostrar la precisión de los principios de la física cuántica en los sistemas biológicos. Creía que las diferencias entre los sistemas biológicos y otros sistemas físicos y químicos están simplemente relacionadas con sus estructuras. Sin embargo, aún así, están hechos de elementos naturales y obedecen todas las leyes de la naturaleza en física y química. Trató de explicar que las diferencias en la construcción de dos sistemas pueden causar una gran diferencia en su función, mientras que los elementos básicos aplicados en ambos sistemas pueden ser los mismos. Esto debe diferenciarse de cualquier otra cosa que pueda causar una comprensión sesgada de los elementos básicos y la naturaleza de un sistema. La diferencia en la construcción es suficiente para hacer una forma de funcionamiento completamente diferente. Schrodinger trató de aclarar este concepto comparando las diferencias estructurales entre un motor térmico y un circuito electrónico, los cuales pueden aplicar los mismos elementos básicos (por ejemplo, cobre y hierro). Se refirió a la consistencia de las propiedades naturales del cobre a pesar de que se ha aplicado de manera diferente en la construcción de un motor térmico en comparación con un circuito electrónico. Para decirlo crudamente, un ingeniero, familiarizado únicamente con los motores térmicos, después de inspeccionar la construcción de un motor electrónico, estará preparado para encontrarlo funcionando según principios que aún no comprende. Encuentra el cobre familiar para él en los hervidores que se utilizan aquí en forma de cables largos enrollados en bobina; el hierro que le es familiar en palancas y barras y cilindros de vapor aquí llenando el interior de esas bobinas de alambre de cobre. Se convencerá de que es el mismo cobre y el mismo hierro, sujeto a las mismas leyes de la naturaleza, y tiene razón en eso. La diferencia en la construcción es suficiente para prepararlo para una forma de funcionamiento completamente diferente. El concepto de coherencia en el tema de la ley de la naturaleza también ha sido mencionado anteriormente por Max Planck y Einstein. Sin embargo, Schrodinger intentó proporcionar una descripción sistemática de la consistencia de las leyes de la naturaleza en los sistemas vivos alineados con la mecánica cuántica.

Schrodinger se refirió al "orden" como la principal diferencia entre un sistema vivo y cualquier otra cosa que exista en la materia inanimada. El desarrollo de eventos en el ciclo de vida de un organismo exhibe una admirable regularidad y orden, incomparable con cualquier cosa que encontremos en la materia inanimada. Lo encontramos controlado por un grupo de átomos sumamente bien ordenado, que representa solo una fracción muy pequeña de la suma total en cada celda.

Además, basándose en el comportamiento de los átomos, Schrodinger clasificó todos los sistemas físicos de la naturaleza en dos categorías principales, incluida la física estadística y la física ordenada (ordenada). Recuerda Schrodinger un pequeño artículo interesante de Max Planck sobre el tema del tipo de derecho dinámico y estadístico (Dynamische und Statistische Gresetzmassigkeit). La distinción es precisamente la que hemos etiquetado aquí como "orden del orden" y "orden del desorden". El objeto de ese artículo era mostrar cómo el interesante tipo estadístico de ley, que controla los eventos a gran escala, se constituye a partir de las leyes dinámicas que supuestamente gobiernan los eventos a pequeña escala, la interacción de los átomos y moléculas individuales. El último tipo está ilustrado por fenómenos mecánicos a gran escala, como el movimiento de los planetas o de un reloj, etc. Parecería que el nuevo principio, el principio de orden a partir del orden, al que hemos señalado con gran solemnidad como la verdadera clave para la comprensión de la vida.

Para el nuevo principio que es genuinamente físico, es, en nuestra opinión, nada más que el principio de la teoría cuántica una y otra vez. Pero no podemos esperar que las "leyes de la física" derivadas de él sean suficientes de inmediato para explicar el comportamiento de la materia viva cuyas características sorprendentes se basan visiblemente en gran medida en el principio de "orden a partir de orden". No esperaríamos que dos mecanismos completamente diferentes produzcan el mismo tipo de ley; no esperaríamos que la llave del pestillo abra también la puerta de su vecino. Por tanto, no debemos desanimarnos por la dificultad de interpretar la vida según las leyes ordinarias de la física. Porque eso es precisamente lo que cabe esperar del conocimiento que hemos adquirido de la estructura de la materia viva. Debemos estar preparados para encontrar un nuevo tipo de leyes físicas que prevalezcan en él. Ahora la gran pregunta es ¿cuál sería el nuevo tipo de leyes físicas que prevalecen en la estructura, función y comportamiento altamente ordenado de los sistemas vivos? Aquí, intentamos explicar los mecanismos que causan el comportamiento ordenado de

moléculas y átomos en sistemas biológicos. Discutiremos cómo el orden a partir de la física del orden puede explicarse por la propiedad única de los sistemas biológicos para la generación de códigos moleculares. Como factor adicional en las leyes de la naturaleza, describimos la capacidad de codificación de moléculas en sistemas biológicos que deben considerarse junto con las propiedades de las moléculas para la producción de energía y en masa. Por lo tanto, determinamos el código como la tercera dimensión de las leyes de la naturaleza que se han ilustrado en las propiedades autónomas de los sistemas biológicos. Al considerar el código como la tercera dimensión de las leyes de la naturaleza en física y química, determinamos una nueva rama de la química llamada química cognitiva que imita la física ordenada en los sistemas biológicos. La química cognitiva integra las leyes de la naturaleza en química, física y matemáticas para la codificación, el almacenamiento de datos y la operación por moléculas. La química cognitiva se puede interpretar en terminología física como la relatividad del código, la energía y la masa.

Módulo 1. Introducción a la biología de sistemas

1.1 ¿Qué es la biología de sistemas?

“A menudo digo que cuando puedes medir de lo que estás hablando, y expresarlo en números, sabes algo al respecto; pero cuando no puedes medirlo, cuando no puedes expresarlo en números, tu conocimiento es de un tipo exiguo e insatisfactorio; puede ser el comienzo del conocimiento, pero apenas en sus pensamientos han avanzado al estado de la ciencia, cualquiera que sea el asunto”. William Thomson (Lord Kelvin) 1883.

Aunque Lord Kelvin no era consciente de los grandes avances que uno puede hacer mirando con los números, su exagerada filosofía cuantitativa centra la atención en los posibles beneficios de las álgebras superior, matricial, binaria, geometría diferencial y la probabilidad bayesiana. Una de las grandes tradiciones intelectuales centra la atención en los posibles beneficios de los sistemas dinámicos y complejos, involucra, a la química, y la física para darle valor a los datos. Ya sea pensando en los datos astronómicos que describen las galaxias, agujeros negros, ondas de gravedad o las conductividades térmicas y eléctricas de los materiales, los números en sí mismos son una parte central del telón de fondo fáctico y conceptual de estos campos. De hecho, a menudo el acto de tratar de explicar por qué los números tienen valor en la biología y su biotecnología, termina siendo un motor de descubrimiento. En nuestra opinión, es un buen momento histórico para hacer un esfuerzo similar para formar estudiante con un nuevo perfil científico, en nanobiología y biología de sistemas. Invitamos al lector que nos sigue, a explorar la ciencia de datos aplicada por investigadores experimentados, que simplemente quieren encontrar los mejores valores para los números de interés dentro de la biología moderna.

El método científico ha estado guiando la investigación biológica durante mucho tiempo. No solo prescribe el orden y los tipos de actividades que dan validez a un estudio científico y un sello de aprobación, sino que también ha moldeado sustancialmente la forma en que pensamos colectivamente sobre el esfuerzo de investigar la naturaleza. El advenimiento de la generación de datos de alto rendimiento, la minería de datos y el modelado computacional avanzado ha puesto en entredicho el estado monolítico, antes indiscutible, del método científico. Por un lado, los nuevos enfoques de una biología mecanicista son claramente exitosos y esperan la misma aceptación que los métodos tradicionales, pero por otro lado, reemplazan gran parte del razonamiento impulsado por hipótesis por la argumentación inductiva⁶⁸, que los filósofos de la ciencia consideran problemática. Intrigados por la enorme riqueza de datos y el poder del aprendizaje automático, algunos científicos incluso han argumentado que las correlaciones significativas dentro de los conjuntos de datos podrían hacer que toda la búsqueda de la causalidad sea sólida⁶⁹. Muchos de estos temas se han debatido apasionadamente durante las últimas dos décadas, a menudo con escaso acuerdo. La combinación de métodos dentro de este espacio sin duda dará forma a nuestro pensamiento sobre la naturaleza, con implicaciones para el diseño experimental, la revisión por pares y la financiación, el intercambio de resultados, la educación, los diagnósticos médicos e incluso las cuestiones de litigio.

Los grafos ocupan un lugar destacado en el razonamiento humano, especialmente en la ciencia. La investigación en ciencias cognitivas ha proporcionado importantes conocimientos sobre las inferencias que ofrecen los grafos y ha revelado diferencias en el razonamiento que son posibles gracias a los diagramas instanciados físicamente y los meramente imaginados⁷⁰. En la práctica científica, los grafos ocupan un lugar destacado tanto en la forma en que los científicos razonan sobre los datos como en la forma en que conceptualizan los mecanismos explicativos. Para identificar patrones en los datos, los científicos a menudo los grafican. Si bien algunos formatos de gráficos, como los gráficos de líneas, se usan ampliamente, los científicos a menudo desarrollan formatos especializados diseñados para revelar tipos específicos de patrones y, con frecuencia, emplean múltiples formatos para presentar los mismos datos, una práctica ilustrada con formatos de grafos desarrollados en biología circadiana. Los científicos cognitivos han revelado el razonamiento espacial y los procesos de búsqueda iterativos que los científicos implementan para comprender los grafos. Al desarrollar explicaciones, los científicos suelen diagramar los mecanismos que consideran

responsables de un fenómeno, una práctica nuevamente ilustrada con grafos de mecanismos circadianos: por ejemplo la red de TOR. La investigación de la ciencia cognitiva ha revelado cómo tales diagramas para comprender un mecanismo genera un fenómeno causal⁷¹.

En términos generales, la biología de sistemas tiene como objetivo capturar la complejidad dinámica de los sistemas vivos mediante la combinación de estrategias matemáticas, computacionales y experimentales⁷². Una cuestión de investigación fundamental es cómo la función biológica surge de las interacciones de procesos cuyas dinámicas no son lineales y están limitadas por la organización del sistema en su conjunto. La investigación en biología de sistemas está impulsada por problemas complejos que requieren soluciones interdisciplinarias, pero existen diferentes puntos de vista sobre los métodos, valores y objetivos más importantes de la biología de sistemas. Algunos académicos enfatizan la integración computacional de big data de múltiples fuentes como un rasgo característico de la biología de sistemas⁷³, mientras que otros enfatizan que la biología de sistemas es una fusión de la teoría de sistemas con la biología⁷⁴. Las diferencias en los puntos de vista teóricos y metodológicos a veces se caracterizan por las diferencias entre la biología de sistemas pragmática y teórica de sistemas: la primera ve la biología de sistemas como una extensión directa de la genómica, la biología molecular y la segunda destaca la necesidad de una teoría de sistemas formal de los sistemas vivos⁷⁵. En ambos casos, sin embargo, los investigadores deben navegar en lo que se llama un espacio de problemas adaptativos donde el conocimiento y los métodos se reconfiguran y combinan continuamente en nuevos métodos, conceptos y modelos híbridos. La combinación de reflexión teórica e innovaciones metodológicas mediadas tecnológicamente hace que la biología de sistemas sea particularmente intrigante desde una perspectiva filosófica y de diseño. Es discutible si la biología de sistemas aporta algo radicalmente nuevo a las ciencias de la vida. La biología de sistemas se ha descrito en términos tan diferentes como un nuevo "paradigma holístico" o "revolución" en biología a ser simplemente una "palabra de moda" utilizada con fines de financiación⁷⁶. Es difícil señalar cambios históricos radicales o "revoluciones", pero la biología de sistemas se desarrolla en un contexto histórico y tecnológico único que ofrece oportunidades nuevas y emocionantes para la

producción de datos, el modelado y también para el desarrollo conceptual. La biología de sistemas combina estrategias tradicionales de investigación biológica con marcos metodológicos y teóricos de diversas disciplinas, incluidas la física, la ingeniería, la informática y las matemáticas. La biología de sistemas tal vez no tenga precedentes en la medida en que se combinan varias disciplinas, y ya es un enfoque diverso y en rápida expansión. Mientras que la etiqueta de biología de sistemas ganó popularidad recientemente, la biología de sistemas tiene muchos precursores importantes y muchos investigadores han estado investigando en biología de sistemas durante varias décadas. Para mencionar algunos ejemplos, Noble desarrolló el primer modelo matemático viable del corazón latiendo a principios de la década de 1960, Davidson y Britten formularon un modelo de red reguladora de genes (GRN) en 1969, y Mesarović y Takahara desarrollaron un marco teórico para sistemas jerárquicos multinivel en las décadas de 1960 y 1970. De manera similar, varios académicos han impulsado enfoques de sistemas para el control dinámico de procesos biológicos durante muchos años antes de que se usara el término biología de sistemas en su contexto moderno. El desarrollo de la biología de sistemas es, por tanto, un proceso largo con diferentes raíces teóricas y metodológicas. Mesarović acuñó el término biología de sistemas ya en 1968 en el Tercer Simposio Internacional de Sistemas en el Case Institute of Technology, Cleveland, Ohio⁷⁷. La biología de sistemas en este contexto inicial supuso el lanzamiento de un campo de investigación biológica basado en la Teoría Matemática de Sistemas Generales, inspirado en el objetivo de la cibernética de identificar y formular principios matemáticos que sustentan capacidades funcionales como el control de retroalimentación. Aún se está explorando el potencial de este marco para el avance conceptual en biología.

Weiss y Bertalanffy son otros precursores cuyo trabajo es de continua relevancia para el desarrollo de teorías en la biología de sistemas contemporánea⁷⁸. De manera similar, destaca cómo muchos aspectos de la biología teórica de Waddington, anticipan las ideas modernas de la biología de sistemas, incluido el énfasis en las propiedades epigenéticas globales y dinámicas como un requisito previo para comprender el desarrollo y la herencia⁷⁹. Otras ramas de la biología de sistemas se inspiran principalmente en el desarrollo del análisis de control metabólico, la teoría de los sistemas bioquímicos o el marco teórico para los sistemas de reparación metabólica⁸⁰. Noble rastrea la historia de la biología de sistemas incluso más atrás hasta Claude Bernard, quien introdujo el principio de homeostasis en 1865. Además, Noble destaca cómo, ya dos siglos antes (en

1665), el filósofo Benedict de Spinoza destacó la importancia de las restricciones a nivel de sistema en los componentes⁸¹. Lo que es común a todos ellos es el reconocimiento de la profunda complejidad de los sistemas vivos, que requieren enfoques que puedan comprender la función biológica en el contexto del sistema como un todo. Otro punto de vista compartido es que puede ser necesaria una teoría formal de sistemas complejos para este propósito. La biología de sistemas desafía el punto de vista, encontrado por muchos modeladores, de que la complejidad biológica es incompatible con los objetivos del análisis matemático formal⁸². Como observa Voit, la opinión de que los sistemas biológicos son demasiado complicados para ser susceptibles de análisis matemático se ha invertido ahora; los sistemas biológicos pueden ser demasiado complejos para comprender sin el uso de matemáticas⁸³. Es importante destacar que el papel mejorado del modelado matemático y computacional en biología también debe entenderse en términos de un contexto histórico y tecnológico de la biología de sistemas que hace que tales herramientas sean una necesidad. Irónicamente, quizás, la culminación del éxito de las estrategias reduccionistas y centradas en los genes en biología resultó en una comprensión cada vez mayor de las limitaciones de los enfoques que investigan componentes o vías moleculares aisladas⁸⁴. Se debe evitar cualquier contraste demasiado simplificado entre la biología molecular y la biología de sistemas, ya que la relación entre ellas es compleja y está lejos de ser clara⁸⁵. Pero a medida que surgieron nuevos conocimientos sobre la complejidad biológica con mejoras para la producción de datos, quedó claro que las herramientas tradicionales de modelado en biología eran insuficientes tanto para manejar la enorme cantidad de datos como para estudiar la dinámica reguladora de redes complejas⁸⁶. La biología de sistemas, en términos generales, implica un “giro cuantitativo⁸⁷” donde los investigadores complementan los estudios de propiedades cualitativas de moléculas específicas con una búsqueda de patrones dinámicos en grandes redes basadas en datos cuantitativos sobre interacciones biológicas. Sin embargo, es importante señalar que, además de este enfoque “de arriba hacia abajo”, otras estrategias en biología de sistemas se describen mejor como de abajo hacia arriba⁸⁸. Las diferencias no son claras, pero los enfoques de arriba hacia abajo tienen como objetivo "aplicar ingeniería inversa" a los patrones regulatorios en grandes conjuntos de datos de tecnologías de alto rendimiento, mientras que los enfoques de abajo hacia arriba o intermedios se basan en datos más detallados (pero a menudo menos completos) de fuentes de biología molecular⁸⁹.

Kitano comenzó a usar el término “biología de sistemas” en este contexto moderno a fines de la década de 1990⁹⁰. Este fue el período que condujo a la finalización del Proyecto Genoma Humano y los primeros proyectos de modelado de big data. Esta noción moderna de biología de sistemas ganó rápidamente vigencia con los desarrollos institucionales desde 2000 en adelante, incluido el surgimiento de conferencias internacionales sobre biología de sistemas (la primera en Tokio en 2000), departamentos de biología de sistemas y revistas dedicadas al trabajo en biología de sistemas. El rápido desarrollo de la biología de sistemas en el siglo XXI muestra que la tecnología no es solo una herramienta en la ciencia, sino que también puede dar lugar a nuevos campos a través de la generación de resultados inesperados y la exploración de un nuevo conjunto de preguntas de investigación⁹¹.

Además, la biología de sistemas ofrece nuevas oportunidades para estrategias de investigación biomédica intensivas en datos, como modelos y simulaciones de múltiples escalas⁹² con profundas implicaciones epistémicas, sociales y éticas⁹³. Las discusiones metodológicas y teóricas en y sobre la biología de sistemas aportan nueva luz sobre muchos temas filosóficos clásicos. Algunos ejemplos son el reduccionismo, la explicación científica, el modelado, la relación entre teoría y experimento, la colaboración interdisciplinaria y el papel de las matemáticas, la ingeniería y la física en la biología. También tienen implicaciones para debates ontológicos más fundamentales sobre qué es la vida y cómo describir los sistemas vivos. Además, surgen nuevas preguntas como las implicaciones del modelado de redes y las simulaciones a gran escala comenzó a usar el término “biología de sistemas” en este contexto moderno a fines de la década de 1990.

1.2 Temas de filosofía de la Biología de Sistemas

1.2.1 Reduccionismo y lo emergente

Como se mencionó anteriormente, la biología de sistemas a menudo se define en oposición a las metodologías reduccionistas. El dicho de que el todo es más que la suma de las partes resalta el interés filosófico y científico en las llamadas propiedades emergentes, es decir, propiedades del sistema que no pueden explicarse o deducirse de las partes solamente⁹⁴. El cliché moderno tiene raíces muy antiguas en la filosofía que se remonta a Aristóteles, pero la biología de sistemas aporta nueva luz a la cuestión de cómo debe entenderse "más" en este contexto. Una pregunta importante desde el

principio ha sido qué tipo de sistemas emergentes apoya la biología⁹⁵. Estas discusiones consideran ontologías sobre la naturaleza de los sistemas vivos, pero también temas epistémicos sobre la metodología apropiada para investigarlos. Numerosos personajes abordan la cuestión de hasta qué punto la biología de sistemas rompe con el reduccionismo y qué significa decir que emergen las propiedades de los sistemas⁹⁶. El enfoque en redes reguladoras de genes, en lugar de secuencias de ADN estáticas como "códigos", es un paso crucial hacia una visión más matizada de las complejidades de las relaciones entre genotipos y fenotipos⁹⁷. La biología de sistemas en este sentido conlleva una epistemología y ontología diferente centrada en la regulación dinámica de los sistemas biológicos⁹⁸. Específicamente, se aboga por una ontología de procesos que también requiere nuevos enfoques metodológicos para alcanzar su potencial. Igualmente importante es si el estudio de los procesos a nivel molecular es suficiente para comprender las propiedades de macroescala, o si se necesita un análisis multinivel⁹⁹. Las contribuciones brindan diferentes puntos de vista sobre la cuestión de si se puede decir que un nivel o escala específicos tienen prioridad causal. Por ejemplo, la investigación sobre la regulación de los procesos de desarrollo debe comenzar con información reguladora genómica en la que las funciones biológicas están "codificadas de forma determinativa", mientras que otros expresan escepticismo sobre la prioridad causal de las redes reguladoras de genes¹⁰⁰. Los ejemplos de modelos de múltiples escalas en el contexto de la investigación sobre simulaciones cardíacas, biología del desarrollo e investigación del cáncer son particularmente esclarecedores para abordar esta cuestión. Es importante ver que algunos investigadores destacan cómo una postura no reductiva también implica una comprensión de cómo los sistemas vivos de alguna manera son menores que la suma de las partes, porque debemos tener en cuenta cómo las restricciones a nivel de sistema influyen en el comportamiento de las partes¹⁰¹. Hofmeyr aclara cómo las actividades de las enzimas están limitadas por el entorno químico y el contexto celular, mientras que Noble destaca la importancia de las limitaciones de escala celular y tisular para las oscilaciones generadoras de pulso responsables de los ritmos cardíacos. Tales ejemplos pueden iluminar discusiones filosóficas sobre los difíciles conceptos de causalidad descendente y control jerárquico. La discusión también tiene implicaciones prácticas. Los debates teóricos sobre si el cáncer es una enfermedad genética causada por mutaciones o refleja un problema de organización de los tejidos también se refieren a cuestiones sobre la metodología más relevante, por ejemplo, si el camino a seguir en la investigación del cáncer es invertir más en la genómica del cáncer o estudiar tejidos *in vitro* e *in silico*¹⁰². Por lo tanto,

razonar sobre los sistemas biológicos de cierta manera puede tener importantes implicaciones prácticas y sociales, y esto es una de las razones por las que la reflexión sobre los supuestos subyacentes de la práctica científica es relevante tanto para los filósofos como para los científicos.

1.2.2 Modelado matemático y computacional en biología de sistemas

Los modelos matemáticos y computacionales se consideran cada vez más indispensables para comprender la complejidad biológica y para integrar e interpretar la gran cantidad de datos de tecnologías de alto rendimiento. En lugar de analizar los detalles de una vía molecular específica, gran parte de la investigación en biología de sistemas se ocupa de la detección de patrones en las arquitecturas de redes que representan interacciones reguladoras interconectadas. Comprender las implicaciones del uso de herramientas teóricas de grafos para el análisis de la organización biológica se ha convertido en un tema importante en la filosofía de la biología de sistemas. El análisis teórico de grafos abarca históricamente desde el análisis de redes aleatorias y retículas regulares hasta el descubrimiento reciente de que muchas redes comparten una estructura de mundo pequeño y libre de escala con importantes implicaciones para sus funciones¹⁰³. La importancia del análisis de redes globales para la investigación biológica es un tema controvertido, y una cuestión importante es si la biología de sistemas puede salvar la brecha entre los enfoques local y global¹⁰⁴. Se destaca un candidato para un camino intermedio, a saber, la búsqueda de patrones de conexiones que ocurren con frecuencia¹⁰⁵. Un ejemplo que ya ha recibido mucha atención de los filósofos de la biología es el trabajo pionero de Alon en el análisis de motivos, es decir, subcircuitos sobreabundantes como los bucles de alimentación¹⁰⁶. La búsqueda de patrones generalizables de circuitos de red también se exhibe en la investigación sobre procesos de desarrollo, sobre redes de reacción bioquímica¹⁰⁷, y sobre la robustez en varios sistemas biológicos¹⁰⁸. Aunque la búsqueda para identificar principios de diseño realizados en circuitos biológicos se remonta a los enfoques teóricos de sistemas mencionados anteriormente, la detección de motivos en redes reguladoras recibió especial atención porque se basó en datos de alto rendimiento y porque las predicciones matemáticas se combinaron con investigaciones experimentales. El resurgimiento del interés por el diseño general o los principios organizativos también ha dado lugar a discusiones sobre la medida en que los sistemas vivos, a pesar de la complejidad de los procesos entrelazados, están constituidos por

unidades funcionales individuales que exhiben modularidad¹⁰⁹. Además, el renovado interés en el análisis matemático de las características organizativas ha llevado a discusiones sobre las implicaciones de la abstracción de los detalles moleculares con el fin de identificar características organizativas generalizables¹¹⁰.

Otro aspecto importante del modelado matemático y computacional en biología de sistemas se refiere a la integración e interpretación de datos en simulaciones a gran escala. Los modelos y simulaciones a gran escala tienen un potencial interesante para las aplicaciones médicas¹¹¹, pero también plantean importantes cuestiones epistémicas sobre (i) la medida en que la complejidad biológica se puede capturar de manera significativa *in silico*; (ii) cómo se combinan los modelos de ingeniería, física y matemáticas, y (iii) el potencial y la necesidad de desarrollar un nuevo marco matemático o lenguaje artificial específicamente adecuado para lidiar con la complejidad biológica. Para hablar de "modelado de células completas" y qué niveles de rendimiento de la predicción del modelo es realista alcanzar. Estas son preguntas abiertas importantes para la investigación actual y futura.

Actualmente estamos siendo testigos de ambiciosos esfuerzos para desarrollar modelos de "pacientes digitales" de órganos completos o completos que abarquen la complejidad biológica de una manera sin precedentes¹¹². Pero es una pregunta abierta hasta qué punto estos desarrollos llevarán la investigación biomédica a destacar cómo los desafíos para simular procesos complejos en meteorología y otras ciencias computacionales son cada vez más relevantes a medida que la biología comienza a desarrollar modelos computacionales complejos. En comparación con los campos computacionales establecidos, se describe la situación actual en biología como el "salvaje oeste" debido a la variedad de enfoques diferentes, lo que hace que muchos modelos sean difíciles de entender y reutilizar¹¹³. Para comprender los resultados de los modelos construidos algorítmicamente, tanto los **filósofos como los biólogos deben esforzarse cada vez más por comprender las técnicas matemáticas involucradas**. Los debates se refieren no solo a qué modelos son los más apropiados para fines específicos y cómo deben combinarse, sino también a si necesitamos un lenguaje digital novedoso para razonar sobre las funciones biológicas. Kitano sostiene que nuestro lenguaje natural es a menudo un obstáculo para la investigación biológica porque es metafórico y sensible al contexto y, por lo tanto, impreciso. En su opinión, la mejor manera de superar el problema es desarrollar enfoques computacionales y un lenguaje artificial exacto y

sistemático. Una de esas iniciativas es el lenguaje de marcado de biología de sistemas¹¹⁴ (SBML), cuyo propósito es servir como una lengua franca legible por máquina que permite la comunicación y traducción entre programas de software. Sin embargo, queda por ver si iniciativas como la SBML conducirán a avances importantes o si ahogarán el progreso y la libertad académica. El modelado a gran escala también plantea el problema de la relación entre el tamaño de los conjuntos de datos y el poder predictivo de los modelos. Algunos colaboradores son optimistas de que conjuntos de datos más completos y potentes simulaciones computacionales pueden superar muchas limitaciones de los modelos actuales e incluso abordar el problema del "cisne negro" de Popper. El "cisne negro" en este contexto es un hallazgo inesperado en la investigación científica, utilizado por Popper (1959) para ilustrar la asimetría lógica entre la verificación y la falsedad de hipótesis científicas.

Independientemente de cuántas observaciones parecieran apoyar la idea sostenida por los europeos hasta el siglo XVII de que todos los cisnes son blancos, solo se necesitó una sola observación de un cisne negro para falsearla. Mientras que la teoría de Popper buscaba fortalecer la metodología científica contra el problema de la inducción, la consecuencia del carácter incompleto de las observaciones empíricas es, en su opinión, que las ideas científicas nunca pueden probarse como verdaderas. Una pregunta intrigante es si la **suposición de incompletitud** puede cambiar con la biología de sistemas, porque el análisis de todo el genoma "ofrece una respuesta impermeable a la preocupación de que es extremadamente difícil o imposible saber si los cisnes negros, es decir, mecanismos cualitativamente diferentes, están al acecho en otros lugares que en las islas de fenómenos elegidos hasta ahora para el análisis causal. Una cuestión relacionada es si los métodos de biología de sistemas rompen con el principio de la navaja de Occam, que establece que se debe preferir la más simple de dos explicaciones. Mientras que Kolodkin y Westerhoff han argumentado que la navaja de Occam no es adecuada para la investigación en biología de sistemas¹¹⁵, sostiene que todas las prácticas de investigación, incluida la práctica de investigación real en biología de sistemas: se basan en una variedad de suposiciones e idealizaciones simplificadoras. Además, otros colaboradores destacan cómo demasiados detalles pueden dificultar la comprensión en la investigación biológica y que parametrizar y validar modelos ricos en datos es un desafío difícil¹¹⁶. Por lo tanto, se debate hasta qué punto los modelos y simulaciones complejos y con gran cantidad de datos nos liberarán de los problemas actuales que surgen de la necesidad de simplificar los espacios de problemas para hacer

que el análisis sea manejable. Mientras que algunos intentos de mejorar los modelos pueden conducir simplemente a una reproducción de la complejidad biológica en modelos poco comprensivos, algunos modelos útiles de múltiples niveles ya han resultado de la investigación biológica de sistemas. Algunos ejemplos son los modelos del corazón virtual que combinan la geometría de los tejidos, la fisiología celular, la expresión génica y el transporte de iones. Estos modelos proporcionan excelentes estudios de caso para el análisis filosófico de los desafíos de la validación de modelos en la práctica¹¹⁷.

La biología de sistemas proporciona así nuevas fuentes para el análisis filosófico de las estrategias de representación para organizar y analizar la información biológica. Mientras que la filosofía de la ciencia se ha centrado tradicionalmente en los textos científicos, las posibilidades y limitaciones de las diferentes formas de representación se están convirtiendo ahora en fuentes ricas para el análisis filosófico del papel cognitivo de las visualizaciones en la biología de sistemas¹¹⁸. Las estrategias de representación en biología de sistemas van más allá de los diagramas de caja y flechas secuenciales e incluyen modelos de redes y diagramas de espacio de fase muy complejos. El uso y el énfasis en estrategias de representación específicas también plantea desafíos para la colaboración interdisciplinaria, ya que los científicos con diferente formación a menudo tienen diferentes puntos de vista sobre lo que es un buen modelo para un propósito específico. Similar, las preferencias por estrategias de representación específicas pueden reflejar diferentes objetivos explicativos en la práctica científica.

1.2.3 Explicaciones y principios de diseño

En lugar de apelar a las leyes, los filósofos de la biología han argumentado que la biología funcional proporciona explicaciones mecanicistas citando cómo las funciones biológicas surgen de la interacción y organización de las partes componentes¹¹⁹. Actualmente existe un animado debate filosófico sobre si las explicaciones de la biología de sistemas son mecanicistas o no. Mientras que gran parte de la investigación en biología de sistemas apoya y produce directamente explicaciones mecanicistas¹²⁰, otros esfuerzos de investigación parecen más difíciles de conciliar con el enfoque mecanicista en partes y operaciones descomponibles y localizables. Por lo tanto, algunos han argumentado que la integración de aspectos de la teoría de sistemas, la ingeniería y la física en las explicaciones de la biología de sistemas revela la necesidad

de explicaciones filosóficas novedosas¹²¹. El análisis matemático de redes ofrece una extensión poderosa de la heurística mecanicista de descomposición funcional y estructural al ofrecer información sobre las características generalizables de cómo se organizan los mecanismos¹²². Un ejemplo destacado es la investigación sobre motivos de red que informa directamente el análisis experimental de circuitos reguladores concretos¹²³. En este contexto, el enfoque intensificado en las características organizacionales generalizables y el modelado dinámico cuantitativo da como resultado actualizaciones y mejoras de las cuentas mecanicistas, pero no una desviación de estas. En respuesta a los desarrollos en biología de sistemas y otros campos de investigación matemáticamente intensivos, algunos filósofos han pedido una explicación mecanicista dinámica actualizada y ampliada¹²⁴. Si bien, la biología de sistemas parece ofrecer nuevas formas de proporcionar explicaciones mecanicistas, otros aspectos de la biología de sistemas pueden describirse mejor como enfoques no mecanicistas¹²⁵. Por ejemplo, algunos académicos han argumentado que algunas prácticas de modelado y explicación se describen mejor apelando a ideales explicativos similares a la ley¹²⁶, explicaciones topológicas¹²⁷, o explicaciones de diseño¹²⁸. La diversidad de ideales explicativos en la biología de sistemas crea nuevos espacios para el análisis científico y filosófico, pero también plantea desafíos para la colaboración interdisciplinaria. Hoy se enfatiza la importancia y las implicaciones de la creciente dependencia de metodologías y marcos teóricos de la física y la ingeniería¹²⁹. Muchos investigadores en biología de sistemas tienen experiencia en ingeniería, y el lenguaje funcional basado en analogías de circuitos y conceptos como ruido, control, amplificadores, filtros, robustez, etc. es cada vez más prominente.

Mientras que este lenguaje es compatible con el lenguaje funcional en biología en general, esta “forma de pensar” en biología de sistemas suele ir acompañada de un creciente interés en identificar características organizativas bastante abstractas que, en general, hacen que un sistema funcional pueda exhibir capacidades específicas. La influencia del control y la teoría de sistemas en el razonamiento en biología de sistemas sugiere que puede ser posible alcanzar una comprensión abstracta a nivel de sistema de los esquemas funcionales básicos sin una comprensión detallada de los mecanismos específicos. La biología de sistemas ha generado un nuevo tema de investigación filosófica, a saber, las implicaciones de los principios de diseño y los principios organizativos y su importancia para la comprensión biológica. Entre las cuestiones relevantes se encuentran no solo si son compatibles con explicaciones mecanicistas o

explicativas por derecho propio, sino también si los filósofos de la biología deberían prestar más atención a otros objetivos científicos que a la explicación. Los candidatos alternativos incluyen predicción, control y diseño¹³⁰. Otra cuestión importante es hasta dónde podemos llegar con una perspectiva de ingeniería en biología. Aunque los principios de diseño y los principios de organización a menudo se usan indistintamente, las preferencias por uno de estos términos también reflejan una preocupación por las posibles limitaciones de un enfoque de ingeniería para los sistemas biológicos. Esta pregunta ha recuperado su relevancia en las discusiones contemporáneas sobre metodologías de investigación en biología de sistemas y sobre la cuestión fundamental de **¿qué es la vida**¹³¹? ¿Hay algo fundamental y distinto en los sistemas vivos? Si es así, ¿se puede capturar en un marco teórico formal? ¿Hasta dónde podemos llegar con los enfoques de ingeniería en biología? ¿Y qué roles juegan las características funcionales y evolutivas en una visión sistémica del organismo de este tipo?

1.2.4 Biología de sistemas funcional y evolutiva

La filosofía de la biología de sistemas enfatizó la relativa autonomía de la biología funcional de la biología evolutiva. Excepto por el interés reciente en el modelado, el análisis experimental y especialmente las explicaciones mecanicistas, gran parte de la filosofía de la biología estuvo hasta hace poco dominada por la biología evolutiva. Nada en biología tiene sentido excepto a la luz de la evolución en el contexto de la biología de sistemas. Ir un paso más allá al sugerir que muchas áreas de la biología de sistemas no tienen sentido sin la luz de la evolución. La resistencia a una interpretación evolutiva que lo abarque todo se debe en parte al escepticismo sobre la teoría etiológica de las funciones (es decir, la función como efectos seleccionados). Las explicaciones sobre si se seleccionó un rasgo y cómo a menudo parecen innecesarias para el análisis funcional. Por lo tanto, la biología de sistemas abre la posibilidad de que el pensamiento de diseño y la explicación del diseño puedan dissociarse de los supuestos evolutivos al menos en algunos contextos¹³². En los últimos años, ha surgido la nueva rama de la biología de sistemas evolutivos, un objetivo importante de la biología de sistemas evolutivos es complementar los estudios evolutivos con un análisis cuantitativo que también mira hacia el futuro al centrarse en cómo la estructura reguladora de las redes biológicas influye en el potencial de cambio evolutivo¹³³. La biología de sistemas evolutivos ha proporcionado otra reformulación más de la máxima de Dobzhansky, a

saber, que nada en biología tiene sentido excepto cuando se cuantifica adecuadamente a la luz de la evolución¹³⁴. La evolución microbiana experimental y las simulaciones evolutivas *in silico* ofrecen nuevas posibilidades para comprender los detalles de cómo han evolucionado determinados procesos regulatorios. La biología de sistemas también puede ayudar a comprender el potencial dinámico de cambio a través de una combinación de datos experimentales y análisis matemático basados en la teoría de sistemas dinámicos¹³⁵. El tiempo mostrará si la biología de sistemas facilitará una "síntesis moderna" nueva o ampliada de la biología evolutiva. Además de la síntesis del marco teórico de la evolución por selección natural y la genética Mendeliana a principios del siglo XX, muchos estudiosos han argumentado que una teoría moderna de la evolución también debería ser capaz de explicar el desarrollo y la evolución de la forma biológica¹³⁶. Las nociones importantes en una síntesis tan extendida son la capacidad de evolución, la robustez, los paisajes adaptativos y epigenéticos y la plasticidad fenotípica¹³⁷. Las herramientas de modelado y simulación de la biología de sistemas parecen particularmente adecuadas para este propósito. Una pregunta relacionada es si la filosofía de la biología de sistemas se involucrará con algunas de las nuevas preguntas planteadas en el contexto de la microbiología. La biología de sistemas también puede reunir campos de investigación que anteriormente estaban divorciados debido a diferencias en los estándares epistémicos. Además de intentar tender un puente entre las ideas de la teoría general de sistemas y los enfoques experimentales, las teorías estructuralistas de la formación de patrones pueden recuperar su relevancia en la biología de sistemas. Según los estructuralistas, muchos patrones observados en la naturaleza no son solo productos de la variación genética aleatoria y la selección natural, sino que son el resultado de restricciones más generales impuestas por la complejidad¹³⁸. Los estructuralistas han enfatizado la necesidad de explicar no solo cómo la selección natural conduce a la preservación y adaptación de las estructuras biológicas, sino también cómo los patrones de la variación, como la orientación espacial y los patrones morfológicos en las plantas, surgen en primer lugar.

Muchos biólogos experimentales han considerado el estructuralismo como un enfoque bastante especulativo del desarrollo y la evolución. Pero dado que ahora es posible realizar ingeniería inversa matemática de la estructura dinámica de algunas redes reguladoras de genes a partir de datos experimentales, la biología de sistemas puede proporcionar nuevos conocimientos sobre cómo la forma y la función se relacionan con estructuras reguladoras específicas e investigar las transiciones evolutivas *in silico*. Una

interesante pregunta abierta para el futuro de la biología de sistemas evolutivos es también si los nuevos enfoques experimentales y computacionales convertirán la biología de sistemas evolutivos en un campo de investigación más predictivo¹³⁹, o si conducirán al descubrimiento de nuevos desafíos de complejidad biológica.

1.2.5 Colaboración interdisciplinaria y educación científica

La integración interdisciplinaria es el nombre del juego en biología de sistemas, pero tender un puente entre las diferencias disciplinarias no es una tarea fácil. Varios colaboradores destacan los difíciles desafíos para la colaboración entre investigadores con diferentes antecedentes educativos. Desde una perspectiva filosófica, tales desafíos pueden ser fuentes importantes de comprensión de los diferentes ideales epistémicos que operan en la ciencia. Las brechas disciplinarias también brindan una oportunidad única para que la filosofía contribuya positivamente a la ciencia y la educación científica. Diagnosticar las fallas y los problemas en las interacciones entre científicos puede ayudar a facilitar la colaboración científica porque las diferencias se pueden aclarar y discutir.

Un problema común al que se enfrentan las prácticas interdisciplinarias de la biología de sistemas es la brecha entre las tradiciones de los experimentalistas y los modeladores. Los estudios empíricos de Nersessian sobre la colaboración entre modeladores y experimentales en biología de sistemas integradores revelan ideas interesantes sobre cómo los investigadores a menudo tienen puntos ciegos para las necesidades, valores o limitaciones del otro campo¹⁴⁰. Por ejemplo, los modeladores piden ciertos tipos de datos que los experimentadores no creen que valgan la pena o ni siquiera que sean posibles de producir. A su vez, los experimentadores pueden no comprender las limitaciones y los potenciales del modelado, o pueden ver el resultado de un modelo como una simple reproducción del hallazgo experimental *in silico*. Una biología de sistemas de problemas muy similar a la siguiente: 'nada en un organismo tiene sentido excepto a la luz de un contexto funcional'. Boogerd va un paso más allá al sugerir que "muchas áreas de la biología de sistemas no tienen sentido sin la luz de la evolución". La resistencia a una interpretación evolutiva que lo abarque todo se debe en parte al escepticismo sobre la teoría etiológica de las funciones (es decir, la función como efectos seleccionados). Las explicaciones sobre si se seleccionó un rasgo y cómo a menudo parecen innecesarias para el análisis funcional. Por lo tanto, la biología de

sistemas abre la posibilidad de que el pensamiento de diseño y la explicación del diseño puedan disociarse de los supuestos evolutivos al menos en algunos contextos¹⁴¹.

Otro aspecto del problema es la falta de comprensión con suficiente detalle de los métodos en los que los científicos en particular no están capacitados. Este problema tiene importantes implicaciones para la educación científica en biología de sistemas. Hohmann sostiene que los estudiantes a menudo están mal preparados para ingresar a un programa de educación superior en biología de sistemas porque la educación previa generalmente está orientada hacia la disciplina, mientras que la biología de sistemas es inherentemente interdisciplinaria y requiere la combinación de diferentes marcos teóricos. Específicamente, la falta de una formación matemática adecuada en los programas educativos para biólogos experimentales dificulta que los estudiantes adopten un enfoque de biología de sistemas. Igualmente importante, modelar sistemas vivos complejos es una tarea desafiante sin entrenamiento experimental o biológico. Los investigadores involucrados en colaboraciones interdisciplinarias necesitan desarrollar experiencia interactiva que permita la comunicación interdisciplinaria y la comprensión conceptual de otras prácticas de investigación¹⁴². La colaboración interdisciplinaria exitosa, además de la experiencia interactiva, implica lo que ella llama **conciencia epistémica de las diferencias en los supuestos y compromisos**, a menudo implícitos, sobre lo que constituye una buena investigación. Una forma de abordar el problema es establecer y mejorar las infraestructuras de investigación. Otra estrategia es hacer que las herramientas de modelado y los tutoriales estén disponibles en línea para que los experimentadores obtengan una mejor comprensión de los procedimientos de modelado. Una iniciativa es un curso de introducción a la biología de sistemas de resolución de problemas para estudiantes diseñada para estructurar los procesos cognitivos necesarios para investigar un problema específico pero complejo en biomedicina. Esto les dará una mejor comprensión de las limitaciones prácticas de las prácticas experimentales y de las cuestiones importantes para los experimentadores. De manera similar, como se indicó anteriormente, los experimentadores podrían mejorar su comprensión de las implicaciones y posibilidades de diferentes técnicas de modelado si tomaran un curso básico en modelado matemático y computacional. Es importante destacar que se parece apoyar la idea de que la educación en biología de sistemas debería volverse interdisciplinaria desde el principio. La compensación entre el alcance transdisciplinario y la experiencia especializada es un tema importante de interés tanto para la educación científica como

para la filosofía. Abordarlo puede requerir preservar algunos límites disciplinarios a pesar del creciente énfasis en la interdisciplinariedad en la ciencia moderna¹⁴³.

1.2.6 Modos de investigación y estructura institucional

La investigación en biología de sistemas requiere nuevas habilidades educativas, pero también conduce a reflexiones sobre los marcos institucionales de la ciencia moderna¹⁴⁴. Varios investigadores destacan que la integración de la biología de sistemas en los sistemas universitarios mediados tecnológicamente plantea importantes desafíos para la innovación científica. Si bien, los investigadores reconocen cada vez más la profundidad y el alcance de la complejidad biológica, continuamente enfrentan limitaciones en términos de oportunidades de financiamiento, métodos disponibles y plazos permitidos para realizar proyectos de investigación. Las propuestas de subvenciones y los comunicados de prensa dan continuamente la impresión de que las innovaciones tecnológicas finalmente nos han posicionado para lograr avances tan esperados en biología y medicina, a pesar de que las metodologías que se persiguen son muy limitadas y no hacen justicia a la complejidad biológica. Jaeger también expresa su frustración con lo que él ve como una tendencia nociva en la ciencia y la sociedad modernas para enfocarse en beneficios inmediatos y cuantificables. Cuando se opera bajo la presión de obtener resultados inmediatos, los proyectos de investigación pueden apuntar a frutos maduros que no resuelvan los problemas conceptuales o sociales más profundos. Dada la complejidad de los sistemas vivos, ignorar las profundas incertidumbres y dificultades asociadas con las metodologías de investigación puede obstaculizar el progreso en el tratamiento de problemas más fundamentales que requieren tiempo para la contemplación y la creatividad. Aunque puede resultar obvio para la comunidad científica que tales afirmaciones son demasiado optimistas, pueden influir en la forma en que se organiza la investigación científica y se asignan los recursos. Si bien estos temas son temas amplios relacionados con la cultura de la ciencia en general, pueden ser particularmente relevantes en relación con la biología de sistemas, porque hay un desarrollo hacia proyectos de “gran ciencia” que involucran a múltiples instituciones internacionales y organismos de financiamiento que invierten fuertemente en ciencia aplicada.

La biología de sistemas brinda oportunidades únicas para desarrollar metodologías radicalmente diferentes, pero también corre el riesgo de desperdiciar recursos en

proyectos que no conducen a ninguna parte. Gross sugiere que los filósofos, como observadores más neutrales, pueden ser capaces de proporcionar una perspectiva útil a "ojo de pájaro" sobre las posibilidades y limitaciones de diferentes enfoques y sobre el desarrollo general del campo¹⁴⁵. Un tema relacionado es la ramificación de las estrategias de uso intensivo de datos que se ha argumentado para cambiar fundamentalmente la forma en que se hace la ciencia. Por un lado, la producción masiva de datos y el aumento de la potencia computacional generan esperanzas de que modelos y conjuntos de datos más completos puedan excluir mejor explicaciones alternativas para un fenómeno dado. Davidson, Drack y Peter también critican la idea de que "los datos pueden hablar por sí mismos". De manera similar, los filósofos se han opuesto a la idea de que el análisis basado en la correlación debería ser un enfoque de investigación más libre de sesgos (Gross 2015). Esta no es una crítica a la investigación intensiva en datos como tal, sino a una práctica que saca grandes conclusiones de datos de baja calidad y que afirma estar libre de sesgos mientras se basa en suposiciones importantes sobre la descomponibilidad modular de las redes biológicas. Si bien, las estrategias intensivas en datos pueden crear nuevas oportunidades de investigación interesantes, es importante discutir si las nuevas estrategias son más eficientes y hasta qué punto deben integrarse mejor con el análisis experimental. Estos problemas también están relacionados con preguntas sobre cómo se producen, curan y utilizan las fuentes de datos en la práctica¹⁴⁶. Los procedimientos de estandarización son importantes, pero también lo es la evaluación de qué tipos de datos biológicos son relevantes para abordar cuestiones específicas. Aunque las ciencias de la vida experimentan actualmente lo que se ha descrito metafóricamente como una avalancha de datos o una avalancha de datos, la falta de disponibilidad de datos de curso temporal de alta calidad sigue siendo uno de los factores limitantes para la investigación en biología de sistemas. De manera similar, aunque hoy tenemos modelos de red tremendamente complejos desarrollados a partir de datos de alto rendimiento, debemos ir más allá de las redes estáticas de "bolas de datos" que no logran captar cómo operan y cambian los procesos biológicos con el tiempo.

1.2.7 ¿La biología de sistemas como el futuro de la medicina?

Se espera que la biología de sistemas juegue un papel central en la medicina del futuro, y los proyectos bajo las etiquetas de medicina de sistemas, medicina personalizada, medicina P4 y medicina de precisión indican las direcciones que podemos esperar que

siga la medicina¹⁴⁷. Al mismo tiempo, todavía no está claro qué significa personalizar la medicina y si las visiones para el futuro son realistas¹⁴⁸. Las simulaciones de múltiples escalas probablemente se convertirán en un tema de interés filosófico cada vez mayor debido a sus profundas implicaciones epistémicas y sociales. Ya se están desarrollando modelos cardíacos de múltiples escalas con implicaciones prácticas para el tratamiento y la elección de fármacos, y actualmente se están tomando medidas para desarrollar un Virtual Physiological Human o modelos digitales de pacientes (Noble 2009). Las simulaciones multiescala pueden tener el potencial de reemplazar algunos estudios experimentales *in vivo* e *in vitro* con grandes implicaciones para la investigación biomédica, los ensayos clínicos y también el bienestar animal. Además, las simulaciones a gran escala desarrolladas en biología de sistemas aportan nueva luz a las discusiones filosóficas sobre la relación entre experimentos y simulaciones. ¿Cuáles son las diferencias y similitudes entre experimentos y simulaciones? ¿Se puede decir que los experimentos por sí solos proporcionan pruebas novedosas en la investigación biológica, o también modelos computacionales? El análisis filosófico del desarrollo y uso de modelos complejos de múltiples escalas muestra que la relación entre experimentos y simulaciones es mucho más compleja y entrelazada de lo que a menudo se supone. La integración de la biología de sistemas con la investigación biomédica y las prácticas de salud también significa que es probable que la filosofía de la biología de sistemas se preocupe más por las cuestiones sociales y éticas. Los problemas involucrarán no solo los conocidos de seguridad de datos, anonimato, confidencialidad y consentimiento, sino también aquellos sobre el cambiante espíritu y valores de la ciencia, como la idea de la medicina de precisión está incrustada en un espíritu comunitario para producir, compartir y estandarizar datos. La organización a gran escala de los esfuerzos biomédicos plantea cuestiones importantes sobre la propiedad y autoría de los datos (Leonelli 2014), sobre lo que significa ser un científico de la vida en estos contextos, pero también sobre el papel que desempeñarán los ciudadanos en la producción de datos en el país. Algunos defensores de la medicina de sistemas han destacado la necesidad de nuevas constelaciones sociales en las que los pacientes y los ciudadanos tengan un papel más activo en la investigación como proveedores de datos de salud (Kumari 2020). Con conjuntos de datos más amplios (p. Ej., De genómica y autocontrol) y potentes modelos de biología de sistemas, la esperanza es que la medicina del futuro pueda predecir y prevenir una serie de enfermedades que hoy en día requieren tratamientos costosos y, a menudo, ineficaces. El enfoque intensificado en la predicción de enfermedades en la medicina de sistemas

plantea interesantes posibilidades para el futuro, pero también preocupaciones sobre la validez clínica y la utilidad de los nuevos modelos, sobre la responsabilidad del paciente y sobre los valores sociales. Las ciencias biológicas juegan un papel crucial en la formación y legitimación de ideologías relativas a los seres humanos, los animales, el medio ambiente y las relaciones entre ellos. Si bien la biología de sistemas puede describirse como emancipadora al alejarse del tipo de reduccionismo perseguido en la biología molecular, el optimismo por la predicción y el control en algunas corrientes de la medicina de sistemas se basa en una imagen simplista de la sociedad humana y ideales deterministas. Si bien un enfoque holístico puede liberar la investigación biomédica de un tipo de reduccionismo (centrado en componentes individuales), el reemplazo puede ser un holismo tecnocientífico que resulte en una creciente medicalización de personas sanas a través de una mayor monitorización de biomarcadores. Por lo tanto, la medicina de sistemas tiene el potencial de cambiar también nuestra visión sobre la salud y la enfermedad, para bien o para mal. Tales cuestiones muestran un terreno fértil para que la filosofía de la biología de sistemas participe más activamente en la configuración de los enfoques en desarrollo de la biología de sistemas y la medicina de sistemas. En última instancia, el desarrollo de la medicina de sistemas brinda una oportunidad única para proyectos de colaboración que involucran a biólogos de sistemas, otros científicos naturales, clínicos, historiadores de la ciencia, científicos sociales y filósofos.

1.3 Perspectivas futuras de la filosofía de la biología de sistemas

La biología de sistemas ha recibido especial atención por parte de los filósofos, porque las cuestiones filosóficas han sido fundamentales para la biología de sistemas desde el principio. Como enfoque inherentemente interdisciplinario, la biología de sistemas reevalúa continuamente los supuestos metodológicos y ontológicos de las ciencias de la vida. A menudo se invita explícitamente a los filósofos a participar en este discurso. En la medida en que la filosofía se ocupa de las reflexiones sobre cómo entendemos el mundo y las implicaciones de suposiciones o visiones del mundo específicas, la conexión entre la filosofía y la ciencia no es sorprendente. Sin embargo, la filosofía de la ciencia estuvo durante muchos años divorciada de la práctica científica y se ocupó principalmente del análisis lógico y conceptual abstracto de las teorías científicas como estructuras de argumentos proposicionales. Como resultado, muchos filósofos genuinamente interesados en la práctica científica han experimentado cierta resistencia

por parte de los científicos practicantes que cuestionan si el análisis filosófico tiene algo que ofrecer a la ciencia. El número de publicaciones conjuntas de filósofos y científicos en ejercicio es una indicación positiva del interés mutuo en la filosofía de la biología de sistemas.

Aunque los filósofos y los científicos en ejercicio suelen estar interesados en las mismas cuestiones, los filósofos suelen abordarlas desde una mayor distancia. En el espacio entre una perspectiva especializada e inevitablemente interesada de los científicos en ejercicio y la distancia reflexiva de la filosofía, se encuentra un potencial único para la reflexión crítica que se basa en cómo se hace realmente la ciencia. Los filósofos pueden comprender los procedimientos experimentales y matemáticos involucrados en la investigación de biología de sistemas, pero los científicos también pueden beneficiarse de la identificación de supuestos, implicaciones y aspectos cargados de valores de la práctica científica que a menudo solo son visibles desde una mayor distancia. La utilidad de la filosofía como una forma de evitar verse abrumado por los detalles científicos que pueden desviar el enfoque de las cuestiones importantes. Al mismo tiempo, pide una participación más activa de los filósofos en la investigación actual. De manera similar, si bien los filósofos a menudo están al margen de las actividades biológicas de los sistemas, también están equipados para un papel de asesores o asistentes. Sin embargo, como señala Carusi, para desempeñar tal papel, la filosofía académica profesional debe reinventarse y estar más informada y comprometida con la práctica científica. El gran potencial de los métodos empíricos como las entrevistas y los estudios de observaciones para informar el análisis filosófico. La colaboración entre la filosofía y la ciencia requiere la voluntad de dejar de lado preocupaciones y motivaciones puramente filosóficas en aras de estar abierto a lo que es importante para la práctica científica. Una experiencia común para los filósofos es que los términos técnicos pueden usarse de formas diferentes y, a menudo, más indeterminadas en la ciencia en comparación con las discusiones filosóficas. Aunque la filosofía puede ayudar a la ciencia en la clarificación conceptual del significado de términos importantes como modelo, validación, mecanismo, etc. en diferentes campos, no está nada claro que la práctica científica requiera que los términos se fijen en definiciones analíticas. Por lo tanto, las colaboraciones entre ciencia y filosofía también deben navegar en espacios de problemas adaptativos que requieren una conciencia epistémica de los diferentes objetivos, concepciones y métodos en la filosofía y la práctica científica. Recién ahora estamos comenzando a comprender lo que las estrategias de investigación y los marcos

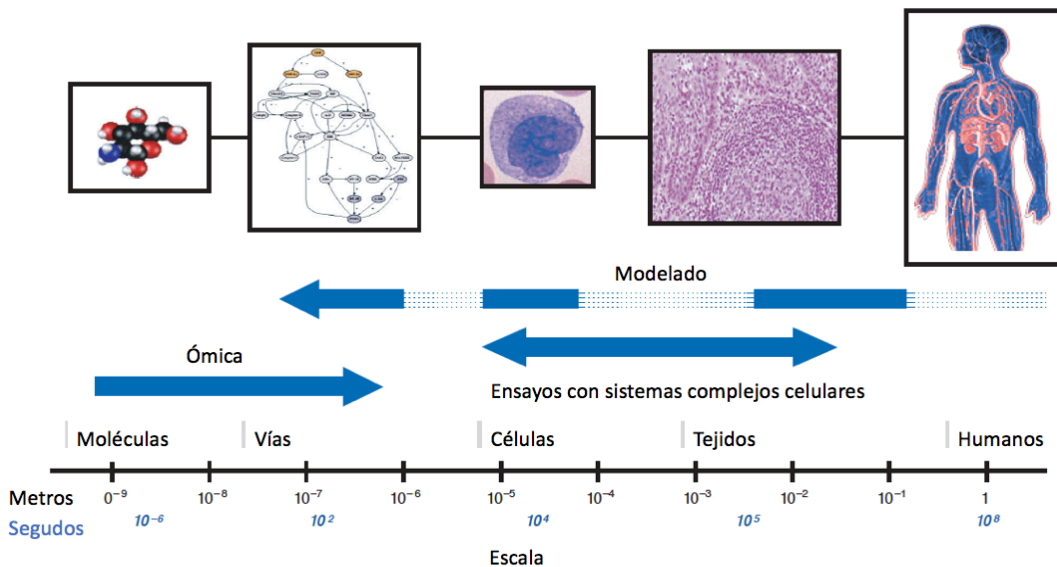
teóricos en biología de sistemas implican para nuestras perspectivas de resolver los grandes enigmas que brinda la complejidad biológica. La biología de sistemas generará alimento para el pensamiento filosófico durante muchos años. Además, la expectativa de que la biología de sistemas tendrá un gran impacto en la investigación biomédica hace que este enfoque sea un tema relevante para cualquier persona interesada en comprender las implicaciones de los desarrollos científicos para la ciencia y la sociedad.

Módulo 2. La biología de sistemas vista panorámica

2.1 Contexto

La vida en un contexto científico amplio se puede definir como el fenómeno que surge de partículas de materia inorgánica organizadas en moléculas que interactúan entre sí dentro de una célula¹⁴⁹. Esta propiedad es sistémica porque solo aparece en el sistema y no en sus partes¹⁵⁰. Los sistemas vivos son estructuras complejas, modulares y jerárquicas. De hecho, un organismo multicelular está formado por moléculas, como ácido desoxirribonucleico (ADN), ácido ribonucleico (ARN), proteínas, lípidos y metabolitos implicados en reacciones químicas y estructuras de las células. Las células se organizan en tejidos que forman órganos con funciones específicas que son necesarias para la salud del organismo. Las propiedades sistémicas aparecen en cada nivel, por ejemplo, homeostasis y respuesta a estímulos en una única red intracelular, metabolismo, crecimiento, adaptación, reproducción en una sola célula. La información que define a un organismo y su capacidad para reaccionar a su entorno está codificada en su ADN y se expresa de forma diferencial en el espacio y el tiempo a lo largo de la vida. Los estudios típicos en biología han utilizado hasta hace poco el enfoque reduccionista y han abordado cuestiones específicas empleando uno o unos pocos tipos de moléculas a pequeña escala, cada una de las cuales arroja luz sobre solo una pequeña fracción de fenómenos enormemente complejos. Algunos hallazgos fueron notables, como el descubrimiento de la estructura del ADN y, más tarde, la forma en que la información genética almacenada en el ADN se transcribe en el ARN mensajero (ARNm) y luego se traduce en proteínas, componentes esenciales de la maquinaria celular y motores de la vida. La acumulación de ese conocimiento sobre moléculas y mecanismos llevó al enfoque "de abajo hacia arriba" para modelar sistemas biológicos, utilizando genes como elementos centrales para simular células, órganos y todo el organismo. Esto era complementario a la visión "de arriba hacia abajo" de un organismo como un sistema fisiológico que integra información de sus diversos componentes y su interacción con el medio ambiente.

Los grandes avances tecnológicos de los últimos 15 años han permitido a los biólogos recopilar información a mayor escala en varios tejidos, incluidas muestras obtenidas con métodos no invasivos, como la recolección de sangre y orina. El aumento masivo del rendimiento ha tenido varias consecuencias. Primero, los biólogos ahora pueden estudiar la gran mayoría de constituyentes, es decir, "algunos", de un elemento dado, por ejemplo, genes de un sistema ya sea un organismo, órgano, célula o para todos los genes de su genoma. En segundo lugar, el gran tamaño de los conjuntos de datos implica que su análisis se basa cada vez más en herramientas computacionales y en el poder disponible para los analistas. En tercer lugar, porque la caracterización de varios "omis", por ejemplo el **genoma, transcriptoma, proteoma y metaboloma**, progresa rápidamente junto con otras disciplinas como la imagen y, en particular, la investigación farmacéutica con quiminformática, bibliotecas de compuestos, cribado de alto rendimiento, seguridad y datos clínicos¹⁵¹, ahora se puede intentar desentrañar las interacciones entre los diferentes elementos de un sistema biológico, o 'interactoma', para comprender su comportamiento a través de varias escalas de manera holística, en salud y enfermedad¹⁵².



Aproximaciones a la biología de sistemas en la industria farmacéutica. Ómicas (el enfoque de abajo hacia arriba) se centra en la identificación y medición global de componentes moleculares. El modelado (el enfoque de arriba hacia abajo) intenta formar modelos integradores (a través de escalas) de la fisiología y la enfermedad humanas, aunque con las tecnologías actuales, dicho modelado se centra en preguntas

relativamente específicas a escalas particulares, por ejemplo, a nivel de vías o de órganos. Un enfoque intermedio, con el potencial de unir los dos, es generar datos de perfiles (p. ej., perfiles de actividad biológicamente multiplexados o datos BioMAP) a partir de ensayos de alto rendimiento diseñados para incorporar complejidad biológica en múltiples niveles: múltiples vías activas que interactúan, múltiples células intercomunicadas tipos y múltiples entornos diferentes. Un enfoque de sistemas celulares tan complejo aborda la necesidad de datos sobre las respuestas celulares a los estímulos fisiológicos y a los agentes farmacéuticos como ayuda para los modeladores, y también como un enfoque práctico para la biología de sistemas en la red de señalización celular y las escalas de interacción célula-célula¹⁵³.

2.1.2 Objetivos y conceptos

La Biología de Sistemas es el estudio integrador de sistemas complejos en la vida con un enfoque holístico ahora basado en conjuntos de datos a gran escala analizados iterativamente con **modelos matemáticos y herramientas de simulación**¹⁵⁴. Comprender cada componente de un sistema complejo de forma aislada no es suficiente para caracterizar el sistema. De hecho, las propiedades del sistema no solo se definen mediante la simple adición de funciones elementales, sino que también surgen de las interacciones entre los elementos¹⁵⁵. Estas propiedades emergentes se estudian infiriendo redes de interacciones entre estos constituyentes, p. ej. genes, proteínas y ligandos, y desentrañando sus mecanismos reguladores. Debido al gran número de elementos en estas redes, tal esfuerzo se basa en conceptos definidos en el marco de la teoría de sistemas complejos¹⁵⁶ y sus límites de reduccionismo. La biología de sistemas no solo tiene como objetivo comprender las relaciones entre los diferentes niveles de expresión de la información genética, a través de la integración de datos, sino también definir el sistema como un todo y producir un modelo matemático convincente, vinculando las interacciones altamente complejas entre sus componentes a sus propiedades emergentes¹⁵⁷. En este contexto, la enfermedad puede verse como un cambio de la homeostasis del rango normal debido a un gran conjunto de perturbaciones en la red de biomoléculas que interactúan en todo el organismo. Por tanto, distintas perturbaciones pueden dar lugar a un único fenotipo de enfermedad, de acuerdo con nuestra comprensión de las enfermedades complejas. Por el contrario, el cambio del sistema a una homeostasis saludable se puede lograr de múltiples maneras y apuntando a varios puntos de la red¹⁵⁸.

La biología de sistemas sigue un enfoque integrador e iterativo que se basa en métodos experimentales y matemáticos (Fig. 1.1). En primer lugar, los datos existentes relacionados con los diferentes niveles jerárquicos del sistema y se integran en modelos matemáticos o gráficos para generar hipótesis, para comprender los mecanismos en juego y construir predicciones sobre las funciones de ese sistema. Luego, algunos componentes del sistema se alteran experimentalmente, como modelos *in vitro* o *in vivo* de una enfermedad. El resultado se evalúa en el contexto del modelo y las hipótesis iniciales se revisan en consecuencia. Estas hipótesis revisadas finalmente informan nuevos experimentos de perturbación. El enfoque se repite hasta que el modelo simula fielmente el comportamiento del sistema¹⁵⁹. Se agrega mayor complejidad cuando se consideran los factores ambientales del modelo.

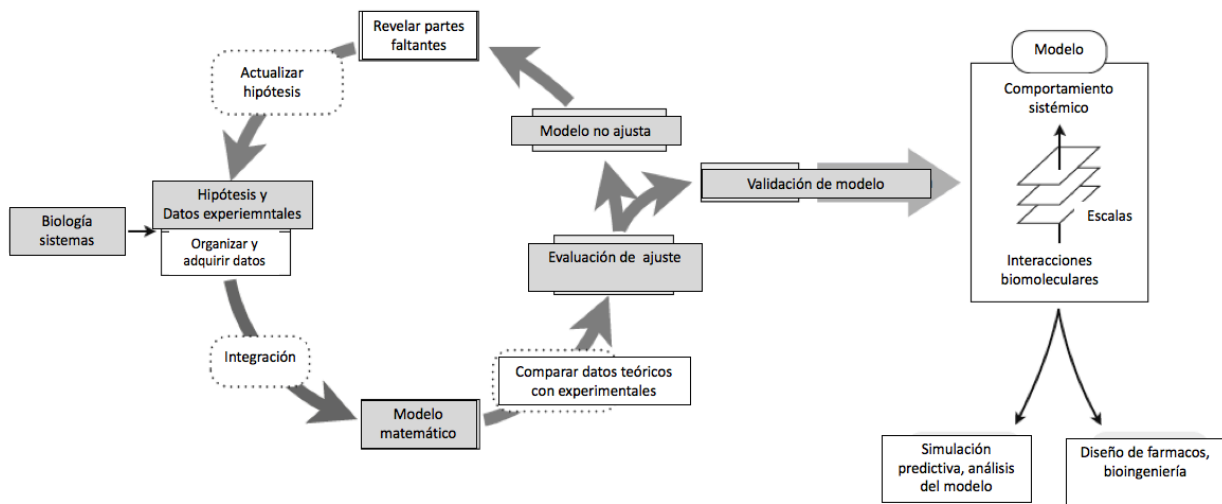


Fig. 2.1 Modelado en biología de sistemas. El modelado comienza con la integración de diferentes datos experimentales en una única base de conocimientos para organizar y almacenar datos. Las descripciones matemáticas de la interacción entre los elementos del modelo permiten (1) la simulación del comportamiento emergente del sistema, (2) la comparación de este comportamiento simulado con datos experimentales, (3) el ajuste del modelo y (4) el diseño de experimentos adicionales. Cuando el modelo se ajusta a los datos experimentales, el estudio del papel de las características particulares del diseño

puede ayudar a identificar los mecanismos en juego y los principios del diseño. El modelo también se puede utilizar en el diseño de fármacos, biotecnología o bioingeniería.

2.1.3 Estrategias

Tres estrategias principales apuntan a construir el vínculo entre los componentes del sistema y sus propiedades emergentes: "de abajo hacia arriba", "de arriba hacia abajo" y "medio hacia afuera" (Fig. 2.2¹⁶⁰). Los pasos principales del enfoque "de abajo hacia arriba" son modelar gráfica o matemáticamente las relaciones entre los componentes del sistema, comenzando con aquellos en el nivel más bajo de la estructura multiescala, por lo tanto, "abajo", p. Ej. genes y proteínas, establezca los parámetros del modelo utilizando valores experimentales y verifique el modelo comparando su comportamiento sistémico con el comportamiento de un sistema real. El término ascendente también se refiere a la dirección elegida: de las propiedades conocidas o supuestas de los componentes se deducen las funciones del sistema¹⁶¹. Esta estrategia de biología molecular ha tenido éxito en el modelado de sistemas biológicos con un número relativamente bajo de componentes, por ejemplo para una sola red intracelular o una sola célula procariota. Sin embargo, puede que no sea adecuado para la reconstrucción de la aparición de sistemas más grandes, por ejemplo, el comportamiento fisiológico de todo el cuerpo en mamíferos. Por el contrario, el enfoque "de arriba hacia abajo" o fisiológico se basa en el comportamiento sistémico. Primero implica definir las formas en que la complicada función sistémica de interés varía con las condiciones y/o el tiempo, y luego inferir las estructuras hipotéticas responsables de esta función. Se perturba el comportamiento del sistema y se estudian los efectos a nivel de los componentes del sistema, es decir, genoma, transcriptoma, proteoma y metaboloma. Esta estrategia está limitada en cierta medida por el desafío de inferir secuencias de ADN a partir de fenotipos. Además, los modelos construidos con enfoques descendentes deben actualizarse con cada nuevo experimento utilizando todos los experimentos existentes, lo que hace que los desafíos analíticos y computacionales sean cada vez más difíciles. Por el contrario, los modelos construidos con el enfoque ascendente, como un modelo de celda *in silico*, comprenden módulos que se actualizan independientemente unos de otros¹⁶². La estrategia "intermedia" pretende superar las limitaciones intrínsecas de los enfoques anteriores, teniendo en cuenta que las cadenas de causalidad pueden operar en los sistemas biológicos en ambas direcciones, comenzando en cualquier nivel de organización biológica. El

comportamiento de un único sistema funcional se modela así en términos de interacciones entre entidades a un nivel suficientemente bien descrito por datos experimentales ("medio"), típicamente de los niveles inferiores de organización, pero no necesariamente hasta las moléculas. Luego, el modelo se extiende a niveles superiores e inferiores ("fuera") de forma iterativa mediante la combinación de enfoques "de abajo hacia arriba" y "de arriba hacia abajo". Se implementó con éxito en el proyecto Physiome¹⁶³. La biología de sistemas jugará un papel crucial en el desarrollo de la medicina personalizada, ya que permitirá la integración de diferentes tipos de datos para perfilar a los pacientes, identificar biomarcadores no sesgados y producir fenotipos de enfermedades precisos. Por tanto, ayudará a la prevención, el diagnóstico y el tratamiento de la Medicina de Sistemas¹⁶⁴.

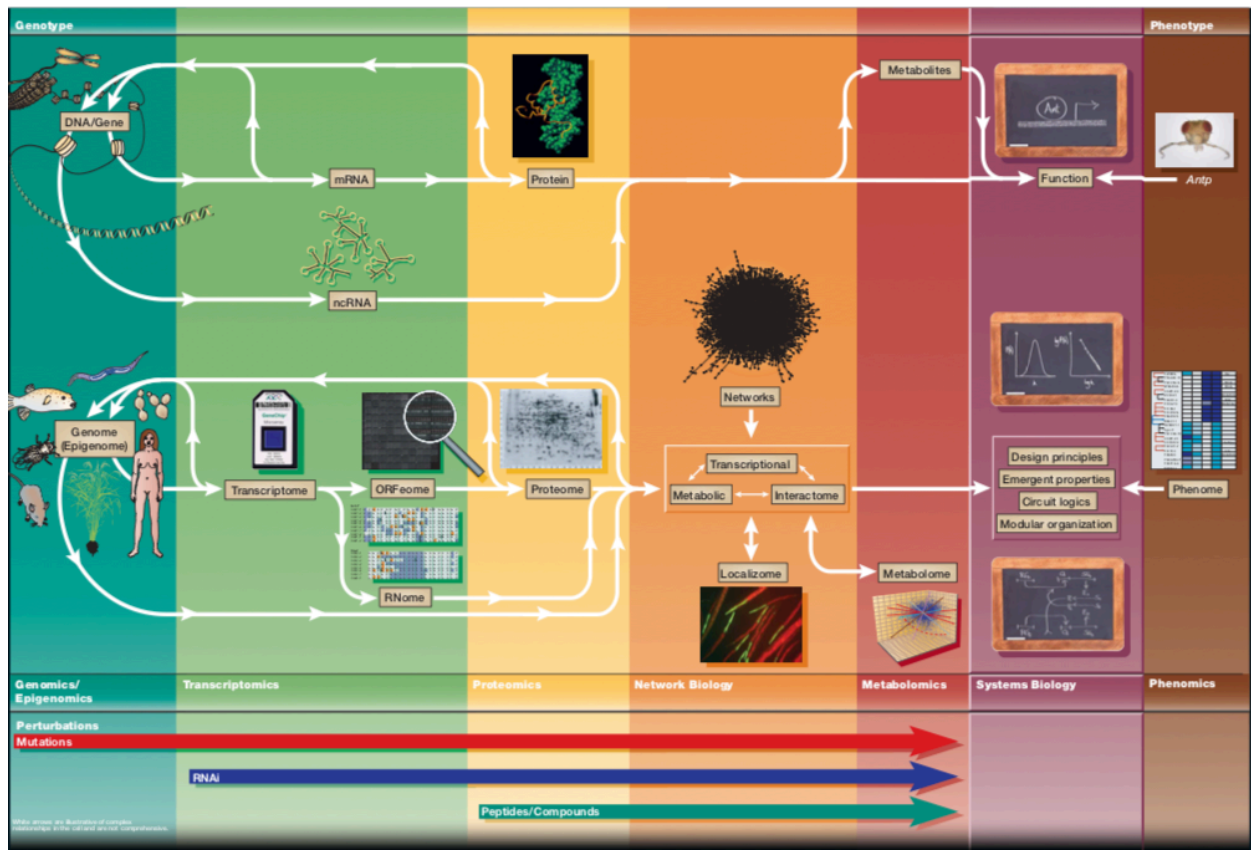


Fig. 1.2 Estrategias de escala múltiple en Biología de Sistemas. Comenzando a nivel molecular, las interacciones entre el ADN, los factores epigenéticos, el ARN, las proteínas, los lípidos y los metabolitos definen los procesos biológicos básicos necesarios para las funciones de orden superior. Estos procesos están definidos por redes de interacción molecular, que se comunican entre sí dentro de una célula determinada, entre células del mismo tejido o tejidos distintos, o entre órganos de un organismo complejo (<http://www.libertadacademica.com/PDFeditorial1/PDF31/ele/>).

2.2 Introducción a la genómica funcional, proteómica, metabolómica y bioinformática

La genómica es el estudio de la secuencia, estructura y contenido del genoma, en particular los genes y su número, estructura, función y organización a lo largo del genoma. La genómica funcional es el estudio de la función de los genes y la regulación de su expresión a nivel de la célula, órgano u organismo, espacialmente y en diferentes momentos y/o estado de salud, mediante el descifrado de la dinámica de transcripción, traducción y proteína de genes. Interacciones de proteínas a escala genómica utilizando tecnologías de alto rendimiento. Las principales herramientas experimentales a gran escala utilizadas para estudiar la epigenética (epigenómica) y la expresión génica (transcriptómica) han incluido hasta ahora microarrays y, más recientemente, la secuenciación de próxima generación. La espectrometría de masas se usa ampliamente para estudiar proteínas (proteómica), metabolitos (metabolómica) y, más recientemente, compuestos orgánicos volátiles (COV) en el condensado del aire exhalado (aliento). Los avances técnicos también llevaron al desarrollo de herramientas computacionales para manejar y analizar su salida.

2.2.1 Tecnologías de secuenciación

La secuenciación del genoma completo comenzó con la secuenciación de un bacteriófago en 1977 utilizando la técnica de secuenciación de Sanger. El desarrollo y la maduración de la secuenciación Sanger automatizada de 4 colores produjo los instrumentos que secuenciaron el genoma humano¹⁶⁵. Posteriormente surgieron varias técnicas de secuenciación de alto rendimiento, o secuenciación de próxima generación (NGS), que eran inferiores a la técnica Sanger automatizada más establecida, siendo más lentas por ejecución, menos precisas, con una longitud de lectura más corta y más

costosa, pero muy superior en virtud del número mucho mayor de nucleótidos leídos¹⁶⁶. Ahora, las estrategias de secuenciación de tercera generación emplean nanoporos y lecturas de una sola molécula, y prometen aumentar el rendimiento y reducir el costo de secuenciación de manera sorprendente. Se están desarrollando herramientas computacionales para procesar la gran cantidad de lecturas cortas y de baja calidad de NGS y ensamblarlas en una secuencia del genoma¹⁶⁷. Las secuencias del genoma de más de sesenta pro y eucariotas se anotan en los navegadores públicos en línea del genoma¹⁶⁸. El conocimiento de genomas completos también permitió el estudio a gran escala de la expresión génica y el desarrollo de la genómica funcional. De hecho, NGS se puede utilizar para análisis de secuencias de ADN o ARN y tiene varias ventajas sobre los microarrays: no requiere un diseño de arreglo, permite estudios de genoma completo a mayor escala, resolución mejorada, más flexibilidad, especificidad de alelo, menor costo y cantidad de material de entrada. NGS ahora también permite el descubrimiento rutinario de variantes en exomas completos e incluso genomas grandes¹⁶⁹ como en Human with the 1000 Genomes Project¹⁷⁰, en la investigación del cáncer¹⁷¹ y estudios de la especificidad de alelos en la expresión génica¹⁷². NGS también catalizó el desarrollo masivo de la metagenómica¹⁷³ y, por lo tanto, las **Máquinas de aprendizaje** ayudarán a descifrar las interacciones huésped-gen-microbiano¹⁷⁴.

Sin embargo, NGS no es lo suficientemente maduro para su uso rutinario en el campo clínico¹⁷⁵. La velocidad, la calidad y la gama cada vez mayores de las aplicaciones de los métodos de secuenciación han creado un enorme flujo de datos y requisitos desafiantes relacionados, no solo para la potencia informática, la memoria y el almacenamiento¹⁷⁶, sino también para el intercambio de datos¹⁷⁷. Las lecturas mapeadas en un genoma de referencia se pueden mostrar con otras fuentes de anotación como NCBI¹⁷⁸ con los navegadores Ensembl¹⁷⁹ y UCSC¹⁸⁰.

2.2.2 Espectrometría de masas

La espectrometría de masas (MS) se basa en la deflexión de átomos cargados por campos magnéticos en el vacío para medir su relación masa/carga (m/z). Un experimento típico sigue cinco pasos: (1) introducción de la muestra, (2) ionización de sus partículas, (3) aceleración, (4) deflexión proporcional a la masa y carga del ion, y

(5) detección, registrada como un espectro que muestra picos en un gráfico de cantidad relativa en función de la relación m/z . Varios métodos de introducción, ionización y tipos de espectrómetros permiten una amplia gama de análisis. Los métodos de introducción son la cromatografía de gases (CG) para mezclas térmicamente estables, la cromatografía líquida (LC) para mezclas térmicamente lábiles y las sondas sólidas. Algunos compuestos, como proteínas grandes y polímeros, deben ionizarse directamente. Los métodos de ionización pueden ser duros o blandos. La ionización fuerte introduce una gran cantidad de energía en las moléculas que da como resultado la fragmentación y, por lo tanto, ayuda a identificar el compuesto, pero los espectros resultantes rara vez contienen el ion molecular. ElectroSpray Ionisation (ESI) utiliza alto voltaje para dispersar e ionizar macromoléculas a través de una boquilla de pulverización. Es suave, limita la fragmentación y produce iones de carga múltiple, lo que permite la detección de compuestos grandes con un valor de masa/carga más baja y, por lo tanto, aumenta el rango de masa del analizador. ESI a menudo se combina con LC/MS. Las mezclas que contienen moléculas no volátiles también se pueden analizar con Fast Atom Bombardment (FAB) y Matrix Assisted Laser Desorption Ionisation (MALDI). MALDI se utiliza para analizar moléculas extremadamente grandes, hasta 200.000 Da, a menudo junto con MS de tiempo de vuelo (ToF). La espectrometría de masas por ionización por desorción láser mejorada en la superficie (SELDI-MS) separa subconjuntos de proteínas fijadas a una superficie de acuerdo con propiedades biofísicas específicas, por ejemplo con hidrofobicidad. Por tanto, el análisis de proteínas, péptidos y nucleótidos se puede realizar con ESI, SELDI, MALDI y FAB¹⁸¹.

Existen varios tipos de analizadores. En un analizador de masas de cuadrupolo (QMS), los iones se desvían mediante campos eléctricos positivos y negativos oscilantes. Un triple QMS contiene tres QMS uno tras otro donde el primer QMS permite la identificación de compuestos conocidos, el segundo en su fragmentación y el tercero en la identificación de los fragmentos, aclarando así la estructura del compuesto. Otros tipos de analizadores incluyen trampa de iones, ToF, Orbitrap y resonancia de ciclotrón de iones por transformada de Fourier (FT-ICR) con una resolución de masa y una precisión crecientes. Los Orbitraps son más baratos, más robustos y tienen un rendimiento más alto que los FT-ICR. La EM en tándem implica varios pasos de selección del compuesto utilizando EM. Los métodos de EM mencionados

anteriormente varían en rendimiento, robustez, sensibilidad, selectividad y facilidad de uso.

2.2.3 Bioinformática

La bioinformática comprende enfoques matemáticos y algoritmos aplicados a la biología y la medicina utilizando herramientas de tecnología de la información, por ejemplo, bases de datos y software de minería¹⁸². El análisis de datos ómicos suele seguir cuatro pasos: (1) procesamiento de datos e identificación de moléculas, (2) análisis de datos estadísticos, (3) análisis de rutas y redes, y (4) modelado de sistemas. Los ejemplos incluyen el ensamblaje del genoma de novo, la anotación del genoma, la identificación de genes coexpresados o expresados diferencialmente a nivel de transcripciones o proteínas y la inferencia de redes de interacción proteína-proteína. La bioinformática también permite la integración de conjuntos de datos heterogéneos de alto rendimiento producidos por un estudio dado y conjuntos de datos existentes utilizando herramientas de gestión del conocimiento, anotación y minería de texto, como los dos vocabularios estructurados, Ontología genética (GO) para genes y procesos biológicos asociados, componentes celulares y funciones moleculares¹⁸³ y la ontología Microarray Gene Expression Data (MGED¹⁸⁴), la base de datos PRoteomics IDentifications (PRIDE¹⁸⁵), Functional Genomics Experiment data model (FuGE¹⁸⁶), Systems Biology Markup Language¹⁸⁷, la notación gráfica de biología de sistemas¹⁸⁸, BioMART¹⁸⁹, tranSMART¹⁹⁰, bioXM¹⁹¹, GARUDA¹⁹², Nexbio¹⁹³ e incluye Biología de sistemas¹⁹⁴.

Estos esfuerzos apuntan colectivamente a desentrañar las vías moleculares que sustentan la fisiología y a identificar biomarcadores para describir un sistema con una combinación de medidas ambientales, clínicas y fisiológicas para mejorar la detección y el seguimiento de un fenómeno, como las enfermedades en la investigación médica, para facilitar el diagnóstico y la terapia. El descubrimiento de biomarcadores se basa en dos tipos de estudios: no sesgados, que solo dependen de la técnica utilizada, y dirigidos, que se centran en biomarcadores predefinidos medidos mediante métodos específicos.

Los métodos y herramientas experimentales y bioinformáticos mencionados en el siguiente texto se enumeran en las Tablas 2.1 y 2.2.

Tabla 2.1 Ejemplos de métodos y herramientas para genómica funcional, proteómica y metabolómica. Esta lista no es exhaustiva.

Métodos de epigenómica: Metilación del ADN: Endonucleasas (MMASS, CHARM, Methyl-seq), conversión de bisulfito (BS) (RRBS, MethylC-seq) y afinidad (MeDIPchip, MeDIP-seq, MDB-seq). A continuación, se pueden medir los niveles de metilación con micromatrices y técnicas de secuenciación; Accesibilidad a la cromatina (DNaseI-seq, FAIRE-seq, Sono-seq, 3C, 4C, 5C, ChIA-PET); Posicionamiento de nucleosomas (CATCH-IT, MNase-se, haploChIP).

Herramientas de epigenómica: Proyecto Enciclopedia de elementos de ADN (ENCODE), el esfuerzo de Epigenomics Roadmap de NIH, el Proyecto de Epigenoma Humano y recientemente BLUEPRINT.

Métodos de transcriptómica: Microarreglo de ADN, SAGE, RNA-seq, ChIP-seq, CLIP-seq.

Herramientas de transcriptómica: ArrayExpress, GEO, MIAME, MINSEQE.

Métodos proteómicos: ELISA, electroforesis en gel 2D, RMN, MS, iTRAQ, SILAC, SRM, SELDIToF herramientas de proteómica MIAPE, TransProteomic.

Métodos metabolómicos: NMR, MS, IMS.

Herramientas de metabolómica: MetabolomeExpress, metaP, KEGG, proyecto del metaboloma humano.

Métodos lipidómicos: MS, orbitraps, IMS.

Herramientas lipidómicas: LIPID MAPS, XCMS, MZmine2.

Tabla 2.2 Ejemplos de métodos y herramientas para bioinformática. Esta lista no es exhaustiva.

Bioinformática: Datos de expresión génica de microarrays (MGED) ontología, base de datos de identificaciones proteómicas (PRIDE), modelo de datos de experimentos de genómica funcional (FuGE), lenguaje de marcado de biología de sistemas, notación gráfica de biología de sistemas, BioMART, tranSMART, bioXM [57], GARUDA [58], nexbio.

Clustering: Babelomics, BASE, MCAM.

Selección de características: Sin supervisión, supervisada; filtros (prueba t de Student, prueba de suma de rangos de Wilcoxon, CFS, EFS, filtrado general de Markov), envoltorios (kNN, Naive Bayes, búsqueda secuencial hacia adelante), métodos híbridos, programación matemática, enfoques de procesamiento de señales.

Análisis de predicción: Sin supervisión (agrupamiento, selección de características, reducción de dimensiones, estimación de densidad y aprendizaje de la estructura del modelo, métodos de reducción de dimensiones no lineales); supervisado (SVM, bosque aleatorio); semisupervisado; serie temporal (HMM).

Redes de literatura: NER, iHOP, FActa +, AliBaba, IntAct, CoPub.

Análisis de ruta o Pathway analysis: Filtrado de expresión diferencial, estadísticas de sobrerrepresentación, GSEA, PAGE, GAGE, ontologizer, GeneCodis, análisis de flujo elemental, vías extremas.

2.3 Genómica funcional, proteómica y metabolómica

2.3.1 Epigenómica

La epigenómica es el estudio de todo el genoma de las modificaciones de la cromatina, es decir, el ADN y las proteínas asociadas, que desempeñan un papel importante en la regulación de los genes, las interacciones entre los genes y el medio ambiente, el desarrollo y enfermedades como la inflamación y el cáncer¹⁹⁵. Tales modificaciones involucran al ADN mismo pero no a su secuencia, es decir, una citosina metilada (mC) adyacente a una guanina (dinucleótidos CpG en mamíferos) y de proteínas de cromatina, es decir, metilación, acetilación y fosforilación de histonas. La epigenómica también cubre la accesibilidad de la cromatina, la remodelación de nucleosomas, las interacciones de cromatina de largo alcance y las firmas de cromatina específicas de alelos. Los avances tecnológicos permiten ahora los estudios de asociación de todo el genoma o EWAS, similar a los estudios de asociación del genoma amplio o GWAS¹⁹⁶, y estudios a gran escala en diferentes tipos de células y tejidos, como en el proyecto ENCYclopedia Of DNA Elements (ENCODE) humano¹⁹⁷, el esfuerzo de NIH Roadmap Epigenomics¹⁹⁸, el Proyecto Epigenoma Humano¹⁹⁹ y recientemente

BLUEPRINT que tiene como objetivo determinar el epigenoma de 100 tipos diferentes de células sanguíneas²⁰⁰.

La metilación del ADN en CpG se estudia ampliamente, ya que media la represión de genes de una manera específica de la célula al evitar que la maquinaria transcripcional acceda al ADN. El ADN metilado se puede detectar con tres tipos de tratamientos de ADN, es decir, endonucleasas, conversión de bisulfito (BS) y afinidad. Los niveles de metilación se pueden medir con micromatrices y técnicas de secuenciación. Las endonucleasas cortan el ADN en sitios específicos, son sensibles a la metilación y permiten varias técnicas de análisis de ADN. Los métodos recientes permiten el análisis de una sola muestra, por ejemplo, evaluación de metilación de muestras individuales basada en microarrays (MMASS²⁰¹), mejores análisis estadísticos y métodos para el diseño de arreglos, por ejemplo con conjunto completo de alto rendimiento para metilación relativa (CHARM²⁰²) y la secuenciación NGS ampliamente utilizada de ADN enriquecido para regiones que contienen CpG (Metil-seq²⁰³).

La conversión de BS modifica la citosina no metilada en CpG en un uracilo y, por lo tanto, transforma una diferencia epigenética en una genética detectable mediante microarrays de ADN específicos de metilación con resolución de un solo nucleótido²⁰⁴. Excepto por mC, el ADN tratado con BS comprende solo tres tipos de bases y, por lo tanto, tiene una complejidad de secuencia y una especificidad de hibridación reducidas. Esto se supera enriqueciendo los segmentos que contienen CpG como en la secuenciación de bisulfito de representación reducida (RRBS) con tratamiento BS y NGS. Las alternativas incluyen la secuenciación de BS de genoma completo, aunque es costosa, y la MetilC-seq ampliamente utilizada, es decir, NGS de ADN tratado con BS. El rendimiento y la cobertura pueden aumentar con la secuenciación de nanoporos que puede secuenciar mC directamente, sin tratamiento con BS²⁰⁵.

La identificación de todo el genoma de los sitios de unión al ADN y las proteínas de unión correspondientes se logra principalmente con el enfoque de inmunoprecipitación de cromatina (ChIP) basado en afinidad mediante el cual proteínas de unión al ADN, por ejemplo las histonas y los factores de transcripción se entrecruzan in vivo en células que luego se lisan. El ADN se fragmenta mediante sonificación, se recupera calentando complejos ADN-proteína y se detecta con microarrays (chip-chip) o NGS (seq-ChIP²⁰⁶). La inmunoprecipitación de ADN

metilado (MeDIP-chip y MeDIP-seq) utiliza un anticuerpo monoclonal contra la citosina metilada para enriquecer el ADN metilado de una sola hebra. Algunas alternativas se basan en cambio en la unión de alta afinidad de un complejo proteico de dominio de unión metil-CpG (MBD) para ADN metilado de doble hebra (por ejemplo, MDB-seq²⁰⁷). Los sitios de unión del factor de transcripción se predicen luego en las secuencias identificadas²⁰⁸. CHIP también se usa ampliamente para estudiar patrones de modificaciones de histonas y modificadores de cromatina²⁰⁹. Puede integrarse a otros conjuntos de datos, como con Segway²¹⁰, lo que ayuda al desarrollo del modelo de cromatina²¹¹. El chip junto con la PCR cuantitativa en tiempo real permite el estudio de la dinámica de las interacciones entre el ADN y las proteínas en células vivas durante varios minutos, y ahora se ha adaptado a la tecnología de microfluidos para reducir el número de células y el tiempo necesario²¹².

En los tres tipos de tratamiento, existen al menos 13 métodos analíticos basados en matrices y 10 seq, cuya elección depende de sus características, la cobertura y resolución requeridas, los tipos de sesgo, precisión y reproducibilidad, y también del número de muestras, calidad de ADN disponible (alta para técnicas de afinidad) y cantidad (alta para técnicas de nucleasa) y, en particular, para métodos basados en matrices: el organismo. Los métodos basados en NGS más utilizados se basan en enfoques de BS (RRBS y MethylC-seq) o de afinidad (MeDIP-seq y MBD seq²¹³). El procesamiento de datos de microarrays se ocupa de los artefactos de imagen y escaneo, la corrección de fondo, la normalización de lotes y arreglos y la corrección del contenido de GC y la densidad de CpG. La proporción de moléculas metiladas a no metiladas para un locus dado es una métrica ampliamente utilizada. Se analiza con herramientas desarrolladas para datos de expresión génica, potencialmente erróneamente ya que se basan en suposiciones violadas por los datos de metilación del ADN, por ejemplo, independencia del número de sitios metilados y no metilados, y similitud de la intensidad de la señal en las muestras²¹⁴. El procesamiento de las lecturas de secuenciación implica el mapeo de las lecturas del genoma de referencia, el recuento y/o el análisis de los datos de bisulfito²¹⁵.

Las regiones genómicas de accesibilidad a la cromatina, es decir, un contenido nucleosómico bajo y una estructura de cromatina abierta, albergan potencialmente secuencias reguladoras y pueden identificarse con un ensayo de hipersensibilidad a la

DNasa I de alto rendimiento (DNaseI-seq también conocido como DHS-seq) [87], aislamiento asistido por formaldehído de reguladores elementos seguidos de secuenciación (FAIRE – seq²¹⁶) y Sono-seq²¹⁷. Y la interacción cromosómica de largo alcance se identifica con captura de conformación cromosómica (3C²¹⁸), 3C en chip (4C²¹⁹), copia de carbón 3C²²⁰ (5C) y se acopla con NGS como en el uso de Hi-C [94] y ChIA PET²²¹. El posicionamiento y la remodelación de los nucleosomas se estudian con CATCH-IT²²² y MNase-seq²²³, mientras que haploChIP identifica perfiles de cromatina específicos de alelos²²⁴, incluidos los SNPS que afectan la expresión génica²²⁵. Los métodos para integrar datos epigenómicos son recientes y se están desarrollando actualmente. Los ejemplos incluyen la integración con datos de expresión génica, utilizando un **modelo empírico de Bayes**²²⁶ y la agrupación de datos de metilación del ADN seguidos de **análisis de regresión no lineal**²²⁷. Las herramientas de visualización pueden mostrar datos sin procesar en todo el genoma como con Circos²²⁸ o la salida del análisis de una manera similar a la utilizada para GWAS, utilizando el valor p log10, pero en dos ejes: prueba de diferencia en el estado de metilación y prueba de diferencia en la expresión génica²²⁹.

2.3.2 Transcriptómica

La transcriptómica es la identificación y cuantificación de todo el genoma de especies de ARN, como ARNm, ARN no codificantes y ARN pequeños, en salud y enfermedad, y en respuesta a estímulos externos. Con los microarrays de ADN, los niveles de expresión génica se miden como la cantidad de ARN en la muestra que coincide con el conjunto de sondas fijadas en el arreglo; Las moléculas de ARN se marcan con fluorescencia y se hibridan en la matriz donde se supone que la intensidad de la señal medida para una sonda determinada es proporcional a la cantidad de ARN unido a ella. Los cambios en los niveles de expresión entre condiciones experimentales o muestras con o sin enfermedad por un lado y la similitud del patrón de expresión con un gen con función conocida por otro lado indican las funciones más probables de los genes. Existen dos repositorios públicos principales para conjuntos de datos de expresión génica: ArrayExpress²³⁰ y Gene Expression Omnibus (GEO²³¹), ambos conformes con las directrices de "Información mínima sobre un experimento de microarrays" (MIAME²³²). Aunque los microarrays son una tecnología establecida y muy utilizada²³³, todavía se están desarrollando métodos de procesamiento y análisis

de datos. Por ejemplo, estudios recientes afirman que los modelos de ruido de fondo basados en la distribución gaussiana para la eficiencia computacional pueden no ser apropiados y los métodos no paramétricos pueden albergar una tasa de falsos positivos más baja²³⁴, mientras que la diferencia promedio ponderada parece ser el mejor método para identificar genes expresados diferencialmente. Existen dos alternativas principales basadas en secuenciación que, a diferencia de los microarrays, no se basan en un conjunto de sondas predefinidas y, por lo tanto, se consideran no sesgadas: el análisis en serie de expresión génica (SAGE) y el transcriptoma NGS de todo el genoma (RNA-seq).

SAGE implica etiquetas de secuenciación que son únicas para cada gen y no están definidas a priori. SAGE se utilizó, por ejemplo, para construir perfiles de expresión de ARN no codificantes largos para 26 tejidos normales y 19 cánceres en humanos²³⁵, arrojando luz sobre su función poco conocida²³⁶. El RNA-seq más reciente proporciona secuencias de transcripciones completas, tiene un ruido de fondo muy bajo, ofrece un rango dinámico muy grande, es altamente preciso y reproducible, permite el descubrimiento de nuevos exones, isoformas y transcripciones RNA-seq ya ha demostrado ser muy prometedor, pero aún no está tan maduro como los microarrays²³⁷. Las transcripciones raras y transitorias hasta ahora no detectadas por los métodos actuales se identificaron recientemente con transcriptómica dirigida mediante captura en una matriz de mosaico seguida de NGS²³⁸. Actualmente, algunos protocolos experimentales pueden introducir sesgos debido a los procesos de amplificación, fragmentación y ligadura²³⁹. El desarrollo de normas y directrices sólidas de control de calidad para microarrays se produjo durante una década, pero debería ser más rápido para RNAseq. Se están desarrollando métodos para describir experimentos que utilizan directrices de 'Información mínima sobre un experimento de secuenciación de alto rendimiento' (MINSEQE) similares a MIAME²⁴⁰, mapear la gran cantidad de secuencias de lectura corta²⁴¹, evaluar los niveles de expresión y detectar transcripciones expresadas diferencialmente²⁴².

Las estimaciones de los niveles de expresión de las transcripciones de interés deben validarse mediante RT-qPCR y técnicas emergentes como la visualización directa y el recuento de moléculas de ARN²⁴³. Sin embargo, estos deben estandarizarse y aplicarse en todas las plataformas. Los microarrays siguen siendo relativamente más baratos que

RNA-seq, sus sesgos son conocidos y los flujos de trabajo de análisis están maduros. Por lo tanto, todavía se prefieren en el descubrimiento de fármacos, aunque los métodos de RNA-seq probablemente los reemplazarán en los próximos años. Dado que los perfiles de expresión génica obtenidos con ambos métodos se correlacionan bien, la gran cantidad de datos adquiridos con microarrays es complementaria a los nuevos datos producidos por RNA-seq²⁴⁴.

Otras técnicas como ChIP también se utilizan para identificar proteínas que se unen al ADN (ChIP-seq) y ARN (CLIP-seq también conocido como HITS-CLIP²⁴⁵). Estos métodos de alto rendimiento de rápida evolución están mejorando en gran medida nuestra comprensión de la regulación de la expresión génica²⁴⁶, en los niveles transcripcional y postranscripcional²⁴⁷.

2.3.3 Proteómica

La correlación entre los niveles de transcripciones y proteínas es incompleta debido a la variación en la velocidad y eficiencia de la traducción y de la degradación del ARNm. Muchas proteínas sufren modificaciones postraduccionales, por ejemplo, fosforilación y ubiquitinación, que modulan su actividad y median en la transducción de señales. Las proteínas también juegan su papel como parte de complejos con otras proteínas o ácidos nucleicos. Un estudio reciente de una línea celular humana identificó más de 10,000 proteínas, con concentraciones que van por encima de siete órdenes de magnitud. Se ha estimado que el proteoma humano comprende varios millones de especies distintas que actualmente no pueden amplificarse y reflejan concentraciones con un rango dinámico muy amplio²⁴⁸.

Las proteínas se pueden identificar utilizando métodos de anticuerpos de bajo rendimiento, ensayos inmunoabsorbentes ligados a enzimas (ELISA) y electroforesis en gel 2D. La proteómica tiene como objetivo definir todas las proteínas presentes en una célula, un tejido o un organismo (o cualquier otro compartimento biológico) y emplea estudios a gran escala y de alto rendimiento del contenido de proteínas, modificaciones, función, estructura, localización e interacciones. utilizando técnicas de alto rendimiento. Los microarrays de proteínas capturan proteínas usando agentes fijados en su superficie, que pueden ser anticuerpos pero también péptidos, receptores, antígenos, ácidos nucleicos. La detección y cuantificación se basan a

menudo en la fluorescencia e identifican interacciones entre proteínas, sustratos de quinasas y activadores de factores de transcripción²⁴⁹. La nanoproteómica tiene el potencial de proporcionar métodos rápidos, sensibles y de alto rendimiento utilizando solo una cantidad mínima de muestras²⁵⁰. Sin embargo, la EM es actualmente la técnica principal para el estudio del proteoma completo a gran escala con mediciones precisas²⁵¹.

La proteómica de escopeta, es decir, LC de escopeta acoplada con MS en tándem (LC - MS/MS) es el enfoque más utilizado. La muestra de péptidos resultante de la digestión de proteínas con tripsina (u otra enzima) se separa mediante cromatografía líquida de alta resolución (HPLC) y los péptidos se identifican mediante EM en tándem: los péptidos se ionizan y separan, produciendo espectros de masas con picos correspondientes a péptidos (primero MS), que luego se identifican mediante una mayor fragmentación y separación de los fragmentos de péptidos resultantes (segundo MS). La inclusión de péptidos sintéticos marcados como muestras de marcado o de adición químicamente (iTRAQ) o metabólicamente (SILAC) mejora la cuantificación²⁵². La complejidad de la mezcla se aborda fraccionando la mezcla. La proteómica dirigida permite identificar de 100 a 200 proteínas en una mezcla compleja identificando previamente los " fragmentos de péptidos de transición " mediante el uso de un espectrómetro de masas de triple cuadrupolo que separa los fragmentos de péptidos de tripsina y luego los fragmenta aún más en " transiciones " que se puede cuantificar en el tercer cuadrupolo. Se intenta elegir transiciones que sean exclusivas de proteínas individuales y la adición de péptidos de transición marcados isotópicamente mejora enormemente la cuantificación. La espectrometría de masas dirigida se denomina monitorización de reacciones seleccionadas (SRM) o monitorización de reacciones múltiples (MRM). Recientemente se han desarrollado ensayos de SRM para todo el proteoma humano (más de 20.000 proteínas) (R. Mortiz, comunicación personal). La HPLC-MS es muy sensible, específica y rápida y, por tanto, se utiliza para el bioanálisis, en particular la farmacocinética para medir la velocidad de eliminación del fármaco por el organismo y en el análisis de muestras de orina. Sin embargo, los inconvenientes incluyen un sesgo hacia la identificación de los péptidos más abundantes. SELDI-ToF es más preciso que el enfoque de escopeta y, por lo tanto, se adapta mejor a la cuantificación de biomarcadores, pero puede no ser lo suficientemente preciso para el diagnóstico clínico²⁵³. Las técnicas recientes producen conjuntos de datos de aproximadamente un millón de espectros, de hasta 100 Gb de tamaño, donde se

pueden identificar hasta 8.000 proteínas²⁵⁴. El preprocesamiento de espectros en bruto implica filtrado de ruido, sustracción de la línea base, detección de picos y calibración y alineación de mapas LC/MS. El análisis sigue cuatro pasos: (1) identificación de secuencias de aminoácidos, péptidos y proteínas en Peptide-Spectrum Match (PSM), y detección, cuantificación, anotación y alineación de características, (2) análisis de importancia de péptidos y proteínas, (3) descubrimiento y predicción de clases, e (4) integración de datos y análisis de rutas. La identificación de secuencias de aminoácidos implica principalmente la búsqueda de bases de datos de espectros obtenidos experimentalmente o de espectros predichos a partir de secuencias genómicas utilizando digestión in silico, y reportando los PSM con las mejores puntuaciones. La fuerza estadística de las predicciones se indica usando la Tasa de Descubrimiento Falso (FDR) calculada usando bases de datos señuelo o modelos que incluyen las proporciones de identificaciones verdaderas y falsas. Debido a que muchos espectros se asignan a muchos péptidos y muchos péptidos se asignan a muchas proteínas, la identificación de péptidos y proteínas es engorrosa y no está completamente resuelta. El problema se complica aún más por las modificaciones postraduccionales y los polimorfismos de un solo aminoácido. Los métodos actuales identifican aproximadamente dos tercios de los espectros de EM en tándem. Las proteínas se informan sobre la base de la coincidencia de un solo péptido, o más estrictamente de la coincidencia con péptidos específicos de proteasa²⁵⁵. Los experimentos se describen utilizando las pautas de Información mínima sobre un experimento de proteómica (MIAPE) similar a MIAME²⁵⁶.

La diferencia en la abundancia de proteínas se evalúa con la cuantificación de proteínas (estimación de la concentración) y la comparación de clases (cambio en la abundancia entre condiciones). El principio es resumir todos los datos cuantitativos relacionados con la proteína mediante (1) recuento espectral, donde se supone que el número de espectros refleja la abundancia con LC MS-MS, y se limita a grandes cambios para proteínas abundantes en mezclas de baja complejidad, o (2) modelos probabilísticos que incorporan todas las características de una proteína y su variación. Estos modelos tienen como objetivo abordar cuestiones importantes, como la representación del diseño experimental, el tratamiento de los datos faltantes y el control de FDR²⁵⁷. Estudios recientes han mostrado ejemplos convincentes de esfuerzos proteómicos cuantitativos realizados en diferentes laboratorios y utilizando varias plataformas experimentales. Actualmente, alrededor de dos tercios de las

proteínas humanas que se prevé que existan se han detectado con EM, de ahí la necesidad de mejorar la sensibilidad, la reproducibilidad de la identificación y la sensibilidad y precisión de la cuantificación²⁵⁸.

Las interacciones proteína-proteína y las cascadas de señalización celular se estudian principalmente con los siguientes enfoques: complementación de dos híbridos de levadura, microarrays de proteínas, cromatografía de inmunoafinidad y MS [137], y con un menor rendimiento por inmunoprecipitación y espectrometría de masas en mamíferos²⁵⁹. Los intentos de integrar proteómica con otros datos ómicos se ven obstaculizados por los inconvenientes actuales del análisis proteómico: proteoma no muestreado completamente, identificación incierta de la proteína, dificultades para mapear identificadores en las diferentes fuentes ómicas, de ahí la necesidad de bases de conocimiento centradas en proteínas como TransProteomic Pipeline²⁶⁰, Protein Atlas y neXProt²⁶¹.

2.3.4 Metabolómica y lipidómica

2.3.4.1 Metabolómica

La metabolómica es la caracterización de alto rendimiento de la mezcla de todos los metabolitos en un sistema biológico, es decir, moléculas pequeñas endógenas y exógenas²⁶². Los metabolitos son lípidos, péptidos y ácidos amino, nucleicos y orgánicos. La metabolómica se usa ahora ampliamente en microbiología, nutrición, agricultura y ciencias ambientales, y en los campos clínico y farmacéutico. Los metabolitos son el producto de reacciones enzimáticas que median procesos biológicos complejos y, por lo tanto, pueden ayudar a comprender los fenotipos. Pueden analizarse mediante espectroscopia de RMN, aunque carece de sensibilidad²⁶³ y la EM (GC y LC) suele preferirse y utilizarse en enfoques dirigidos y no dirigidos. Las estrategias dirigidas son específicas y sensibles, permiten una cuantificación absoluta y, por lo tanto, se utilizan ampliamente en el diagnóstico clínico y el desarrollo de fármacos. Los enfoques dirigidos basados en isótopos estables y modelos de redes metabólicas permiten estimar el flujo a través de vías bioquímicas²⁶⁴. En contraste, los enfoques no dirigidos albergan una alta cobertura, aunque cualquier identificación de metabolitos es menos específica y sensible, y requiere un análisis computacional más

intensivo. Las características que se utilizarán para la identificación se detectan mediante análisis univariados y multivariados y luego se utilizan para buscar en bases de datos como la Enciclopedia de genes y genomas de Kyoto (KEGG²⁶⁵).

A menudo se requieren más experimentos para distinguir isómeros y caracterizar metabolitos no identificados utilizando EM en tándem o RMN. La metabolómica también incluye la identificación del sustrato en ensayos *in vitro* de tres tipos: (1) la proteína se fija a una superficie y se examinan los ligandos, (2) el metabolito se fija y sirve como cebo para las proteínas que interactúan, o (3) se basa en la actividad perfilado de proteínas utilizando sondas químicas y perlas. Por último, pero no menos importante, la ubicación de los metabolitos dentro de las células, tejidos o cuerpos se puede estudiar mediante el acoplamiento de MALDI o MS sin matriz y técnicas de imagen (espectrometría de masas de imágenes, IMS) para obtener espectros escaneando la muestra biológica con el láser y luego compilando un mapa del contenido de metabolitos en esa muestra²⁶⁶.

Se están desarrollando activamente normas para la descripción de experimentos y herramientas para el procesamiento y análisis de datos metabolómicos²⁶⁷. Por ejemplo, MetabolomeExpress y metaP combinan herramientas desde el procesamiento de datos brutos, es decir, la detección de picos de EM, hasta el análisis multivariado. El desarrollo de biomarcadores con metabolómica y la comparación entre conjuntos de datos dependen de: (1) la caracterización de los artefactos técnicos de la EM y las diferencias en los compuestos que discriminan muestras entre analizadores y (2) el tipo de muestra y la variabilidad biológica²⁶⁸. El Proyecto del Metaboloma Humano cuantificó más de 4.000 metabolitos en hasta 70 muestras de las 6.826 identificadas por Wishart et al²⁶⁹. Otro estudio reciente de metabolómica dirigida a gran escala cuantificó 122 metabolitos en 377 individuos, incluidos pacientes y controles con diabetes tipo 2, e identificó 25 metabolitos en plasma y 15 más en suero con concentraciones diferentes en los dos grupos²⁷⁰.

2.3.4.2 Lipidómica

Los lípidos desempeñan funciones importantes en la señalización implicada en el metabolismo, el almacenamiento de energía y la proliferación, migración y apoptosis celular²⁷¹. También son los componentes principales de las membranas celulares, junto

con las proteínas de membrana. De este modo, mantienen la arquitectura celular y median el tráfico de membranas al permitir el ensamblaje de la maquinaria de proteínas, como por ejemplo en grupos dinámicos que recogen proteínas específicas en balsas de lípidos²⁷². Los lípidos son muy diversos en su estructura, propiedades físicas y cantidad. Por ejemplo, los lípidos de señalización y estructurales se encuentran respectivamente en abundancia baja y alta. Los lipidomas, los lípidos presentes en las estructuras biológicas, son poco conocidos en la actualidad²⁷³. El lipidoma humano puede contener miles de especies²⁷⁴, mientras que solo el 20% de todos los lípidos pueden haber sido detectables con las tecnologías existentes²⁷⁵. Los estudios de lipidómica tienen como objetivo caracterizar el contenido, la localización y la actividad de los lípidos en células y tejidos²⁷⁶. La gran mayoría de los lípidos se extraen de células y tejidos lisados, y se analizan con MS ya sea directamente en el método de escopeta, es decir, lipidómica 'de arriba hacia abajo' con analizadores de alta resolución como Orbitraps, o con LC-MS / MS «lipidómica para distinguir lípidos con idéntica relación de carga a masa²⁷⁷. Los lípidos también se han analizado con MALDI IMS²⁷⁸. Los datos brutos de Lipidomics MS pueden analizarse con herramientas utilizadas para metabolómica, como XCMS²⁷⁹ y MZmine 2²⁸⁰.

Los lípidos se identifican y cuantifican mediante el procesamiento de datos sin procesar y el análisis estadístico, seguidos de un análisis de rutas y un modelo²⁸¹. Las principales iniciativas lipidómicas incluyen la "Estrategia de vías y metabolitos lipídicos" (LIPID MAPS), que ha establecido estándares y ha permitido una cuantificación absoluta en lugar de relativa²⁸², y el lipidoma de macrófagos de ratón²⁸³. Las cantidades absolutas para proteómica y lipidómica ayudarán a caracterizar complejos que comprenden tanto proteínas como lípidos.

Los futuros avances técnicos deben apuntar a una mayor precisión, mayor coherencia y armonización de los protocolos. Los desarrollos analíticos deben incluir: (1) procesamiento automatizado de datos e identificación y extracción de lípidos, (2) análisis de datos estadísticos para abordar el cálculo de alta dimensionalidad e independiente de la plataforma de la tasa de descubrimiento falso de identificación de lípidos, (3) análisis de vías para identificar bioquímicos, señalización y procesos reguladores que involucran a los lípidos de interés caracterizados en un conjunto de

muestras, y (4) modelado en tiempo y espacio dentro del contexto de fisiología y sistemas²⁸⁴.

2.4 Métodos y herramientas

Las tecnologías actuales de alto rendimiento producen conjuntos de datos muy grandes y han desplazado el cuello de botella de la producción de datos al análisis de datos. Por lo tanto, las herramientas de gestión del conocimiento son muy valiosas para organizar, almacenar y analizar datos, ya sea directamente con software integrado o indirectamente exportando los datos en el formato requerido. Los conjuntos de datos recientes también albergan una dimensionalidad muy alta. La integración de datos tiene como objetivo combinar conjuntos de datos tan grandes y de alta dimensionalidad que difieren en el tipo de datos recopilados. La integración no supervisada tiene como objetivo reducir la dimensionalidad de grandes conjuntos de datos, sin introducir un sesgo inherente a conocimientos e hipótesis previos. Ayuda a detectar patrones dentro y entre conjuntos de datos y complementa las observaciones estándar en la construcción de hipótesis. Luego, estos se prueban analíticamente con métodos supervisados, usualmente solo usando una fracción de las dimensiones disponibles y experimentalmente²⁸⁵. A pesar de su poder y sus promesas, la integración de datos es solo un medio para un fin, no un motor automático para generar hallazgos valiosos. De hecho, las respuestas a las preguntas formuladas en un estudio científico dependen directamente del diseño experimental, por ejemplo, los tipos de datos, controles, procesamiento y análisis, y el tamaño de las muestras, dentro de las limitaciones financieras y de tiempo. La siguiente sección describe métodos para agrupamiento, selección de características, análisis de predicción, minería de texto y análisis de rutas (Fig. 2.3).

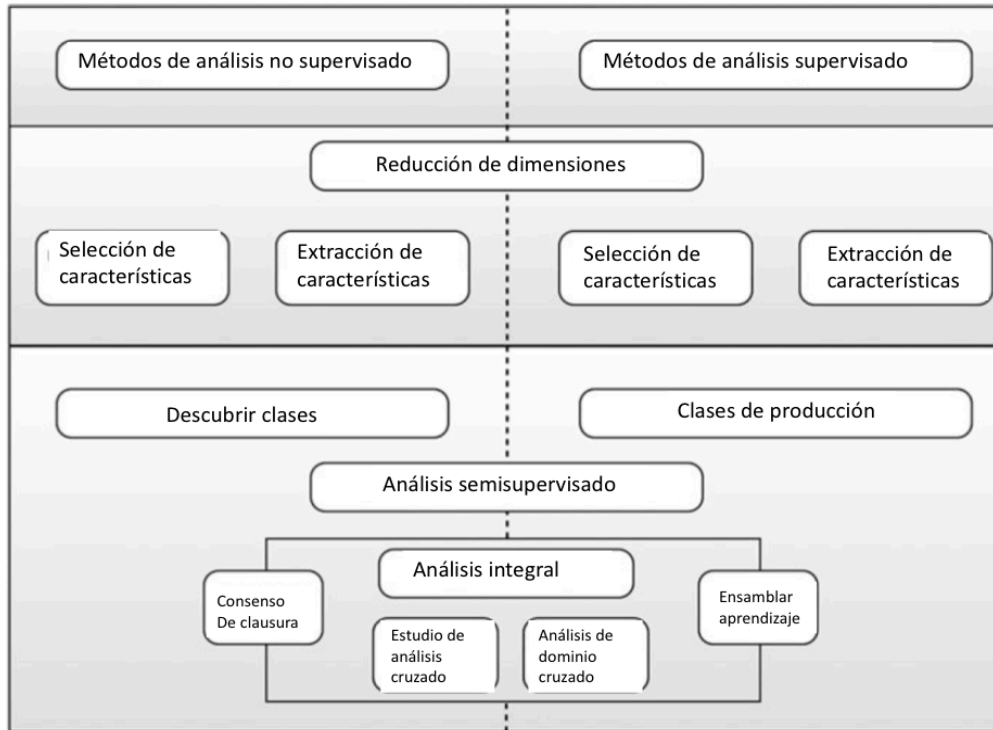


Fig. 2.3 Descripción general de los métodos de aprendizaje automático. Los métodos supervisados y no supervisados van desde enfoques de reducción de dimensionalidad de nivel inferior hasta técnicas analíticas de nivel superior y sus extensiones para el análisis integrador de datos²⁸⁶.

2.4.1 Agrupación

Motivación: la agrupación en clústeres es una técnica de exploración de datos para análisis multivariante que divide los datos en función de grupos intrínsecos sin etiquetas predefinidas. Los métodos de agrupamiento se han aplicado a varios aspectos de la investigación biomédica, por ejemplo, expresión génica en el cáncer, para distinguir pacientes o subgrupos de genes en función de los niveles de expresión de un conjunto de genes expresados diferencialmente. Los genes agrupados pueden tener funciones similares, estar involucrados en el mismo proceso celular o en vías similares.

Tal conocimiento mejoraría nuestra comprensión de la función de los genes y los procesos biológicos. Los métodos de agrupación se pueden utilizar para visualización, generación de hipótesis y selección de genes para análisis posteriores. Procesamiento previo: la agrupación en clústeres requiere métodos de normalización estándar para

datos ómicos²⁸⁷. La agrupación requiere específicamente una reducción de dimensionalidad previa y una estandarización de datos, por ejemplo, filtrar genes o proteínas con baja varianza entre las muestras, métodos basados en la maximización de una función de covarianzas como en el método de "suma de covarianzas"²⁸⁸ (SUMCOV), y estandarización de los datos, por ejemplo, estandarización de la desviación absoluta media. Estado de la técnica: se han desarrollado numerosas herramientas de agrupación. Varios algoritmos de agrupación en clústeres conocidos son: agrupación jerárquica, agrupación basada en particiones y densidad y agrupación difusa. Los algoritmos de agrupación desarrollados más recientemente incluyen: métodos subespaciales o bi-agrupación que agrupan tanto genes como muestras²⁸⁹. La adquisición automática, el preprocesamiento y el análisis de agrupamiento a través de herramientas basadas en web es posible para varias tecnologías de alto rendimiento, por ejemplo Babelomics²⁹⁰, entorno de software BioArray (BASE²⁹¹) y metodología de análisis de agrupamiento múltiple (MCAM²⁹²). Los procedimientos eficientes de validación de grupos son cruciales para la toma de decisiones con una gran cantidad de genes en ausencia de una gran cantidad de muestras y, por lo tanto, serán extremadamente útiles para comprender las interacciones genéticas y diseñar objetivos de fármacos. Casos de uso: la agrupación en clústeres se utiliza ampliamente en el análisis de datos de microarrays y existe una amplia variedad de herramientas. La agrupación de genes puede identificar un grupo de genes con funciones similares, mientras que la agrupación de muestras puede sugerir subgrupos de pacientes para estratificación, respuesta a tratamientos y subtipos o grado de enfermedad, para cáncer de Colon²⁹³, cáncer de mama²⁹⁴ y asma²⁹⁵. Los conglomerados también se pueden integrar con el análisis de rutas²⁹⁶.

2.4.2 Selección de funciones

Motivación: los métodos de selección de características o atributos tienen una amplia gama de aplicaciones en biología de sistemas. Permiten al experimentador identificar qué genes o proteínas se expresan significativamente diferencialmente en distintas condiciones biológicas en un tipo de célula de interés, y qué subconjuntos de genes o proteínas proporcionan el conjunto combinado de biomarcadores más prometedor para discriminar entre estas condiciones. Además, los enfoques de selección de características, se utilizan a menudo para reducir la dimensión de los datos de entrada

antes de aplicar otros métodos de análisis estadístico de nivel superior. Esto alivia una variedad de problemas estadísticos referidos como la maldición de la dimensionalidad en la literatura²⁹⁷. Sin embargo, a diferencia de los métodos de reducción de dimensiones basados en la transformación de características²⁹⁸, las características originales de los datos se conservan, lo que facilita la interpretación de los datos en análisis posteriores.

Los algoritmos de selección de características se pueden agrupar en enfoques supervisados²⁹⁹ y no supervisados³⁰⁰, dependiendo de si incorporan información de etiquetas de clase para las condiciones biológicas. Además, los algoritmos de selección de características que emplean métodos de predicción para puntuar la informatividad de un subconjunto de características se conocen como envoltorios, mientras que otros enfoques univariados y combinatorios para filtrar atributos se denominan filtros³⁰¹. Preprocesamiento: para la mayoría de las plataformas experimentales utilizadas en Biología de sistemas, se requieren varios pasos de preprocesamiento de bajo nivel antes de aplicar métodos de selección de características. Estos incluyen procesamiento de imágenes³⁰², normalización³⁰³ y enfoques de resumen³⁰⁴, para datos de expresión génica de microarrays³⁰⁵ y filtrado de datos brutos³⁰⁶, detección de picos³⁰⁷, alineación de picos y métodos de normalización del tiempo de retención para proteómica y metabolómica para datos de espectrometría³⁰⁸. Además, algunos métodos de selección de características requieren una discretización previa de los datos, por ejemplo, si se utilizan medidas de asociación especiales, como la información mutua³⁰⁹.

Estado de la técnica: la elección del método de selección de características depende tanto del objetivo del análisis (por ejemplo identificar biomarcadores individuales o construir un modelo predictivo combinatorio para la clasificación de muestras) como de la compensación deseada entre la eficiencia (la ejecución de la complejidad temporal del algoritmo) y precisión (el poder predictivo de las características seleccionadas). Entre los enfoques de filtro, los estadísticos univariados simples como la prueba t de Student paramétrica y la prueba de suma de rangos de Wilcoxon no paramétrica todavía se utilizan ampliamente, debido a sus ventajas en términos de velocidad y la dificultad de estimar dependencias de características a partir de datos ruidosos de alta dimensión. Métodos combinatorios más complejos como CFS³¹⁰, EFS³¹¹ y filtrado general de Markov han ganado influencia recientemente³¹².

Los métodos de envoltura son cada vez más populares. Califican subconjuntos de características utilizando métodos de predicción en combinación con un enfoque de exploración espacial de búsqueda y sus selecciones alcanzan el rendimiento predictivo de vanguardia en problemas de clasificación biológica. Los ejemplos incluyen combinaciones de métodos de predicción rápidos y simples³¹³, por ejemplo kNN y Naïve Bayes, y métodos de búsqueda de exploración espacial, como la búsqueda secuencial hacia adelante³¹⁴. Estos enfoques están siendo reemplazados gradualmente por combinaciones de algoritmos más complejos, incluidos algoritmos evolutivos³¹⁵ y métodos de aprendizaje automático basados en el núcleo³¹⁶. Finalmente, varias técnicas recientes han mejorado el equilibrio entre velocidad y precisión: (1) combinación de filtros³¹⁷, (2) combinación de filtros y envoltorios en métodos híbridos, (3) programación matemática³¹⁸ y (4) enfoques de procesamiento de señales³¹⁹. Casos de uso: La identificación y priorización de biomarcadores de genes, proteínas o metabolitos mediante técnicas de selección de características tiene tres objetivos principales: (1) distinguir las condiciones biológicas, por ejemplo, la presencia de cáncer, de infección viral o grados tumorales, (2) median el diagnóstico temprano, la terapia personalizada del paciente, el monitoreo de la progresión de la enfermedad y (3) ayudan a estudiar el tratamiento en un cultivo celular o modelo animal. Sin embargo, los métodos de selección de características también se utilizan para filtrar conjuntos de datos antes de la aplicación de otros métodos de análisis de datos de nivel superior, otros métodos de aprendizaje automático, análisis de sobrerrepresentación de rutas y análisis de redes. Finalmente, la selección de características a menudo se integra con técnicas de clasificación y regresión para disminuir la complejidad de los modelos de aprendizaje automático y maximizar su precisión predictiva.

2.4.3 Análisis de predicción

Motivación: el análisis de predicción se refiere a una familia de métodos que intentan capturar dependencias estadísticas y extraer patrones de un conjunto de datos medidos, para hacer predicciones sobre datos futuros. Estos métodos son muy prometedores en genómica funcional, proteómica, metabolómica y bioinformática, donde las tecnologías recientes proporcionan una gran cantidad de datos, como mediciones de expresión de genes y proteínas, lecturas de secuencias de ADN y ARN. La velocidad a la que se producen estos datos hace que el análisis de predicción

automática sea una herramienta indispensable para el biólogo. Los métodos para el análisis de predicciones pueden ser sin supervisión, semi-supervisados o supervisados. Estado del arte: los métodos no supervisados encuentran regularidades y estructuras ocultas en los datos. Los enfoques típicos incluyen agrupamiento, selección de características, reducción de dimensiones, estimación de densidad y aprendizaje de la estructura del modelo³²⁰. Los métodos clásicos de reducción de dimensiones lineales son el análisis de componentes principales y el análisis de componentes independientes, pero recientemente han aparecido algunos métodos de reducción de dimensiones no lineales muy poderosos³²¹.

Los métodos supervisados utilizan datos en forma de pares (x, y) y estiman una función que predice el valor de y a partir de una entrada x dada. Cuando y es una cantidad discreta (por ejemplo, una etiqueta de varias condiciones biológicas distintas), el método se llama clasificación y cuando y es continuo, el método se llama regresión. El desafío clave es garantizar que la función estimada pueda generalizarse bien en situaciones invisibles³²². Dos métodos son populares: (1) máquina de vectores de soporte (SVM) que estima una función discriminativa maximizando el margen de separación de clases³²³ y (2) bosque aleatorio, basado en conjuntos de árboles y votaciones³²⁴.

Los métodos semi-supervisados combinan ideas de métodos supervisados y no supervisados, para capturar la estructura no supervisada en los datos con el fin de mejorar el rendimiento de la clasificación³²⁵. Los métodos de series de tiempo usan datos medidos en diferentes momentos para modelar y predecir valores futuros de los datos, capturando su estructura y regularidades y contabilizando los efectos estocásticos, como los con modelos ocultos de Markov (HMM³²⁶). Casos de uso: un ejemplo típico es la clasificación de datos biológicos, como datos de expresión génica, en diferentes clases biológicas, por ejemplo enfermo y saludable, principalmente utilizando SVM y bosques aleatorios. Los métodos de predicción también se aplican al análisis de rutas, la descomposición de redes y la anotación de secuencias. A menudo se combinan con una selección de características para extraer las dimensiones más relevantes en el espacio de datos de entrada³²⁷.

2.4.4 Construyendo redes y caminos a partir de la literatura

Motivación: La mayoría de textos une esfuerzos con las ciencias experimentales para ayudar a la investigación multifacética relacionada con enfermedades. Las redes y los mapas de conectividad se derivan del texto en un intento de encontrar conexiones y relaciones causales entre componentes de sistemas biomédicos complejos, con el fin de dilucidar los mecanismos de la enfermedad y detectar comorbilidades³²⁸.

Preprocesamiento: la preparación de datos textuales consiste en la tokenización, eliminación de signos de puntuación, etiquetado de parte de la voz y, a veces, análisis sintáctico. A continuación, se identifican en el texto los nombres de proteínas, genes, sustancias químicas, fenotipos y enfermedades. El manejo de la terminología biomédica aborda varias cuestiones, como la aparición de nuevos términos, el uso intensivo de acrónimos, abreviaturas y palabras de propósito general que designan genes³²⁹. La sinonimia y la homonimia imponen desafíos especiales al proceso de reconocimiento y complican la vinculación de un gen nombre a su identificador único en la base de datos³³⁰. Los sistemas de reconocimiento de identidad (NER) de última generación alcanzan una medida F de aproximadamente el 86³³¹% en el corpus biomédico, frente al 93% en los textos en inglés de propósito general.

Estado de la técnica: la reconstrucción de las vías biológicas a partir de la literatura ha evolucionado desde co-ocurrencias de proteína-proteína por pares no dirigidas³³² a eventos biomédicos complejos de interacciones tipificadas y por lo tanto dirigidas que abarcan múltiples proteínas. Estos últimos se basan en gran medida en los corpus ricamente comentados, el análisis sintáctico profundo y las técnicas de aprendizaje automático supervisado. Debido a la complejidad del lenguaje natural, la extracción precisa de eventos biomédicos sigue siendo un desafío. La medida F lograda por los sistemas de vanguardia varía aproximadamente entre el 70 y el 48%, dependiendo en gran medida del tipo de evento que se reconoce. Casos de uso: muchas herramientas biomédicas de minería de textos ayudan a los usuarios en diferentes etapas del procesamiento de texto, en particular para la construcción de redes y vías. El modelo de co-ocurrencia se ha implementado con éxito en iHop, una red hipervinculada de genes y proteínas mencionadas en los resúmenes de PubMed³³³. Facta amplía el modelo de co-ocurrencia por pares con la extracción de eventos y el descubrimiento de asociaciones indirectas entre los conceptos biomédicos. Basándose en resúmenes de PubMed, AliBaba construye redes de proteínas, genes, asociaciones de enfermedades y

ubicación subcelular de proteínas que interactúan³³⁴. Las redes extraídas de texto se pueden complementar con datos experimentales utilizando IntAct³³⁵ y CoPub³³⁶.

2.4.5 Análisis de ruta

Motivación: el análisis de vías tiene como objetivo identificar las desregulaciones de vías para mejorar la comprensión de fenotipos complejos al aprovechar la información sobre interacciones biomoleculares conocidas en las vías para guiar la búsqueda a través del espacio de posibles asociaciones funcionales. Existe una amplia gama de métodos, que incluyen estadísticas de análisis de enriquecimiento, métodos de priorización de genes de enfermedades basados en vías, análisis de vías metabólicas convexas y métodos de predicción/reconstrucción de vías *in silico*³³⁷.

Procesamiento previo: debido a que las plataformas de medición experimentales y las bases de datos de rutas tienden a usar diferentes formatos de identificadores, el análisis de rutas generalmente comienza con la conversión de los nombres de genes/proteínas a un formato estándar³³⁸, seguido de la normalización y el procesamiento previo de los datos experimentales.

Estado de la técnica: recientemente se han desarrollado varios enfoques novedosos para inferir cambios en la actividad de la vía a partir de datos de alto rendimiento con mayor precisión que mediante la combinación clásica de filtrado de expresión diferencial con estadísticas de sobrerrepresentación como la prueba exacta de Fisher (para conjuntos de datos desordenados) o la prueba de Kolmogorov-Smirnov (para conjuntos de datos clasificados). Estos incluyen enfoques paramétricos y no paramétricos que tienen en cuenta las mediciones del nivel de expresión génica sin filtrar, por ejemplo GSEA³³⁹, PaGE, GAGE³⁴⁰ o explotar información de gráficos de ontología, como Ontologizer y GeneCodis³⁴¹. Para el estudio de las vías metabólicas, dos enfoques relacionados que utilizan el análisis convexo se han vuelto cada vez más importantes: los modos de flujo elemental³⁴² y las vías extremas³⁴³. Finalmente, a diferencia de la definición clásica de vías basada en expertos humanos, recientemente se han propuesto varios métodos para la predicción/reconstrucción de vías utilizando datos experimentales³⁴⁴.

Casos de uso: los análisis de la vía de todo el genoma han proporcionado nuevos conocimientos sobre la etiología de enfermedades complejas que no pueden obtenerse

de los análisis clásicos de un solo locus³⁴⁵. De hecho, tales análisis han demostrado que diferentes alteraciones en una vía pueden causar la misma enfermedad, como en el cáncer colorrectal³⁴⁶. El análisis de vías metabólicas se utiliza en aplicaciones biomédicas y biotecnológicas, por ejemplo para aumentar el rendimiento de producción de microorganismos mediante ingeniería metabólica, es decir, la modificación de vías seleccionadas mediante tecnologías de ADN recombinante³⁴⁷. El análisis de rutas también se puede integrar con el análisis de redes para identificar módulos de redes desregulados en enfermedades complejas³⁴⁸.

2.5 Conclusiones

El estudio de genes individuales y sus productos en sistemas modelo se ha desplazado hacia estudios de alto rendimiento en laboratorios y, a menudo, generados por grandes consorcios. Cada tipo de datos ómicos está resultando muy valioso y su integración promete recompensas aún mayores. Las técnicas actuales son muy diversas y pueden analizar muestras biológicas complejas. Albergan una alta sensibilidad y especificidad, aunque no siempre suficiente, como en la proteómica. Los desarrollos en curso aumentarán la precisión, solidez y flexibilidad al tiempo que reducirán los costos. Las innovaciones técnicas actuales continúan desplazando el cuello de botella de la producción de datos al análisis de datos. Nuestra comprensión de la biología dependerá cada vez más de la gestión de datos y conocimientos, y de la infraestructura informática para complementar los avances en el modelado matemático y computacional para técnicas analíticas temporales y espaciales, que son cruciales para la biología de sistemas.

Referencias

¹ Szilárd, K.. (2013). Against reductionist theories of causation. 57. 142-157.

Mckitrick, Jennifer. (2009). Dispositions, Causes, and Reduction.

² David, Charles. (2007). Working with Max Delbrück. 10.1142/9789812775818_0009.

³ Liu, Wei & Duan, Huaichuan & Zhang, Derong & Zhang, Xun & Luo, Qing & Xie, Tao & Yan, Hailian & Peng, Lianxin & Hu, Yichen & Liang, Li & Zhao, Gang & Xie, Zhenjian & Hu, Jianping. (2021). Concepts and Application of DNA Origami and DNA Self-Assembly: A Systematic Review. Applied Bionics and Biomechanics. 2021. 1-15. 10.1155/2021/9112407.

⁴ O'connor, D.. (1959). The Uses of Argument. By Stephen Edelston Toulmin. Philosophy. 34. 10.1017/S0031819100037220.

⁵ Gonçalves-Segundo, Paulo. (2020). The Functional Configuration of Epistemic Argumentation: Reframing Toulmin's Layout of Arguments in a Multidisciplinary Perspective. Bakhtiniana Revista de Estudos do Discurso. 15. 236-266. 10.1590/2176-457347130.

⁶ Ruyant, Quentin. (2021). Theories, Models and Representation. 10.1007/978-3-030-72349-1_2.

⁷ Nevo, Aviv & Whinston, Michael. (2010). Taking the Dogma out of Econometrics: Structural Modeling and Credible Inference. Journal of Economic Perspectives. 24. 69-82. 10.2139/ssrn.1559385.

⁸ Esarey, Justin & Sumner, Jane. (2017). Marginal Effects in Interaction Models: Determining and Controlling the False Positive Rate. Comparative Political Studies. 51. 001041401773008. 10.1177/0010414017730080.

⁹ Fürstenwerth, Hauke. (2013). On Correlation and Causality. Critical care medicine. 41. e140-1. 10.1097/CCM.0b013e31828c2517.

¹⁰ Cox, Gary. (1999). The Empirical Content of Rational Choice Theory: A Reply to Green and Shapiro. Journal of Theoretical Politics - J THEOR POLIT. 11. 147-169. 10.1177/0951692899011002001.

¹¹ Fernández-Fernández, Claudia. (2021). Formal Model for Explicit Aware Knowledge. 10.1007/978-3-030-69606-1_6.

- ¹² Thicke, Michael. (2020). Evaluating Formal Models of Science. *Journal for General Philosophy of Science*. 51. 10.1007/s10838-018-9440-1.
- ¹³ Grim, Patrick & Singer, Daniel. (2021). Computational Philosophy - Stanford Encyclopedia of Philosophy.
- ¹⁴ Šešelja, Dunja. (2018). Exploring Scientific Inquiry via Agent-Based Modeling.
- ¹⁵ Eberle, Oliver & Buttner, Jochen & Krautli, Florian & Müller, Klaus-Robert & Valleriani, Matteo & Montavon, Gregoire. (2020). Building and Interpreting Deep Similarity Models. *IEEE Transactions on Pattern Analysis and Machine Intelligence*. PP. 1-1. 10.1109/TPAMI.2020.3020738.
- ¹⁶ Borim-de-Souza, Rafael & Travis, Eric & Munck, Luciano & Galleli, Bárbara. (2020). An objective hermeneutic approach to qualitative validation. *Qualitative Research in Organizations and Management An International Journal*. ahead-of-print. 10.1108/QROM-08-2019-1806.
- ¹⁷ Shepherd, Bryan & Jarrett, Ryan & Fu, Lingjun. (2016). GUIDO IMBENS, DONALD RUBIN, Causal Inference for Statistics, Social, and Biomedical Sciences: An Introduction. New York: Cambridge University Press.: Book Review. *Biometrics*. 72. 1387-1388. 10.1111/biom.12615.
- ¹⁸ O. Wolkenhauer et al. Systems biologists seek fuller integration of systems biology approaches in new cancer research programs. *Cancer Res.*, 70(1):12–13, 2010. doi: 10.1158/0008-5472.CAN-09-2676.
- ¹⁹ Alison, Laurence & Smith, Matthew & Eastman, Oliver & Rainbow, Lee. (2003). Toulmin's philosophy of argument and its relevance to Offender Profiling. *Psychology Crime & Law - PSYCHOL CRIME LAW*. 9. 173-183. 10.1080/1068316031000116265.
- ²⁰ Karmakar, Rajesh & Das, Amit. (2021). Effect of transcription reinitiation in stochastic gene expression. *Journal of Statistical Mechanics: Theory and Experiment*. 2021. 033502. 10.1088/1742-5468/abdeb1.
- ²¹ Veliz-Cuba, Alan & Josic, Kresimir. (2013). Piecewise Linear and Boolean Models of Chemical Reaction Networks. *Bulletin of mathematical biology*. 76. 10.1007/s11538-014-0040-x.
- ²² Delehanty, Megan. (2021). Extending the mechanism: Dealing with the context objection to reduction.
- ²³ Wimsatt, William. (1976). Complexity and Organization. *PSA 1972*. 1972. 67-86. 10.1007/978-94-010-1829-6_8.

- ²⁴ Schaffner, Kenneth. (1994). *Discovery and Explanation in Biology and Medicine*. Bibliovault OAI Repository, the University of Chicago Press.
- ²⁵ Bechtel, William & Richardson, Robert. (2010). *Discovering Complexity: Decomposition and Localization as Strategies in Scientific Research*. 10.7551/mitpress/8328.001.0001.
- ²⁶ Glennan, Stuart. (2021). *Mechanisms, models, and causation*
- ²⁷ Cartwright, Nancy. (1994). *Nature's Capacities and Their Measurement*. 10.1093/0198235070.001.0001.
- ²⁸ Salmon, Wesley. (2020). *Scientific Explanation and the Causal Structure of the World*. 10.2307/j.ctv173f2gh.
- ²⁹ Jones, Jude. (1997). *Process Metaphysics: An Introduction to Process Philosophy*. *International Philosophical Quarterly*. 37. 250-252. 10.5840/ipq199737220.
- ³⁰ *Peripheral Nervous Systems: Development, Structure, and Function*. 10.1007/978-3-030-53363-2_22.
- ³¹ Bugajak, Grzegorz. (2011). *Causality and Determinism in Modern Physics*.
- ³² Johnson, R. & Russo, Federica & Schoonenboom, Judith. (2017). *Causation in Mixed Methods Research: The Meeting of Philosophy, Science, and Practice*. *Journal of Mixed Methods Research*. 13. 155868981771961. 10.1177/1558689817719610.
- ³³ Papachristos, George. (2018). *A mechanism based transition research methodology: Bridging analytical approaches*. *Futures*. 98. 57-71. 10.1016/j.futures.2018.02.006.
- ³⁴ Santis, Marcelo. (2020). *Scientific explanation and systematics*. *Systematics and Biodiversity*. 19. 10.1080/14772000.2020.1844339.
- ³⁵ Radder, Hans. (2021). *Empiricism Must, but Cannot, Presuppose Real Causation*. *Journal for General Philosophy of Science*. 10.1007/s10838-021-09551-9.
- ³⁶ Craver, Carl. (2020). *Remembering: Epistemic and Empirical*. *Review of Philosophy and Psychology*. 11. 10.1007/s13164-020-00469-7.
- ³⁷ Siponen, Mikko & Klaavuniemi, Tuula. (2020). *Why is the hypothetico-deductive (H-D) method in information systems not an H-D method?*. *Information and Organization*. 30. 100287. 10.1016/j.infoandorg.2020.100287.
- ³⁸ Prakash, Kirti. (2016). *The periodic and dynamic structure of chromatin*. 10.11588/heidok.00021825.
- ³⁹ Zhang, Xian-En. (2020). *Nanobiology—Symphony of bioscience and nanoscience*. *Science China Life Sciences*. 63. 10.1007/s11427-020-1741-y.

- ⁴⁰ Moreno, Alvaro. (2018). On Minimal Autonomous Agency: Natural and Artificial. *Complex Systems*. 27. 289-313. 10.25088/ComplexSystems.27.3.289.
- ⁴¹ Cox, Michael & Nelson, David. (2000). *Lehninger Principles of Biochemistry*. 10.1007/978-3-662-08289-8.
- ⁴² Kärcher, B. & Hirschberg, M. & Fabian, P.. (1996). Small-scale chemical evolution of aircraft exhaust species at cruising altitudes. *Journal of Geophysical Research*. 1011. 15169-15190. 10.1029/96JD01059.
- ⁴³ Hakim, Md. (2017). Biophysical Modeling of Dietary Energy in Biochemical Modeling. *European Journal of Biophysics*. 5. 57. 10.11648/j.ejb.20170503.12.
- ⁴⁴ England, Dr & Cohen, Simon. (2021). *A User's Guide to Intellectual Property in Life Sciences*. 10.5040/9781526511782.
- ⁴⁵ Zheng, Yao-Rong & Yang, Hai-Bo & Ghosh, Koushik & Zhao, Liang & Stang, Peter. (2009). Multicomponent Supramolecular Systems: Self-Organization in Coordination-Driven Self-Assembly. *Chemistry (Weinheim an der Bergstrasse, Germany)*. 15. 7203-14. 10.1002/chem.200900230.
- ⁴⁶ Jiang, Zongliang & Logsdon, Deirdre & Schoolcraft, W.B. & Krisher, Rebecca & Yuan, Ye. (2018). Single cell gene expression profiling of trophoblast generated from self-organizing human embryos in vitro. *Fertility and Sterility*. 110. e81. 10.1016/j.fertnstert.2018.07.245.
- ⁴⁷ Wang, Xu-Qing & Li, Wei-Jian & Yang, Hai-Bo. (2019). CHAPTER 3: Self-organization in Coordination-driven Self-assembled Metallomacrocycles. 10.1039/9781788013123-00047.
- ⁴⁸ Ding, Zhili & Xiong, Yunfeng & Zheng, Jinxian & Zhou, Dongsheng & Kong, Youqin & Qi, Changle & Liu, Yan & Ye, Jinyun & Limbu, Samwel. (2021). Modulation of growth, antioxidant status, hepatopancreas morphology, and carbohydrate metabolism mediated by alpha-lipoic acid in juvenile freshwater prawns *Macrobrachium nipponense* under two dietary carbohydrate levels. *Aquaculture*. 546. 737314. 10.1016/j.aquaculture.2021.737314.
- ⁴⁹ Espinosa-Diez, Cristina & Miguel, Verónica & Mennerich, Daniela & Kietzmann, Thomas & Sánchez, Patricia & Cadenas, Susana & Lamas, Santiago. (2015). Antioxidant responses and cellular adjustments to oxidative stress. *Redox biology*. 6. 183-197. 10.1016/j.redox.2015.07.008.
- ⁵⁰ Al-Khalili, Jim & Lilliu, Samuele. (2020). *Quantum Biology*. *Scientific Video Protocols*. 1. 1-4. 10.32386/scivpro.000020.

- ⁵¹ Hauer, Jürgen & Kleinekathöfer, Ulrich & Jansen, Thomas & Mancal, Tomas & Miller, R. & Ogilvie, Jennifer & Prokhorenko, Valentyn & Renger, Thomas & Tan, Howe-Siang & Tempelaar, Roel & Thorwart, Michael & Thyryhaug, Erling & Westenhoff, Sebastian & Zigmantas, Donatas & Duan, Hong-Guang. (2020). Quantum biology revisited. *Science Advances*. 6. eaaz4888. 10.1126/sciadv.aaz4888.
- ⁵² MacCord, Kate. (2020). *Cell and Molecular Biology*. *The Quarterly Review of Biology*. 95. 163-163. 10.1086/709075.
- ⁵³ Wang, Jie & Wang, Xin & Fan, Xinyuan & Chen, Peng. (2021). Unleashing the Power of Bond Cleavage Chemistry in Living Systems. *ACS Central Science*. XXXX. 10.1021/acscentsci.1c00124.
- ⁵⁴ Slack, J.M.W.. (2020). *Molecular biology of the cell*. 10.1016/B978-0-12-818422-6.00005-8.
- ⁵⁵ Perez Lopez, Ana & Rubio-Ruiz, Belén & Sebastian, Victor & Hamilton, Lloyd & Adam, Catherine & Bray, Thomas & Irusta, Silvia & Brennan, Paul & Lloyd-Jones, Guy & Sieger, Dirk & Santamaría, J. & Unciti-Broceta, Asier. (2017). Gold-Triggered Uncaging Chemistry in Living Systems. *Angewandte Chemie (International ed. in English)*. 56. 10.1002/anie.201705609.
- ⁵⁶ Qureshi, Muhammad. (2018). *HARPER'S ILLUSTRATED BIOCHEMISTRY 30th*.
- ⁵⁷ <https://cieumich.mx/ContenidoQuantum2021/pages/Quantum21.html>
- ⁵⁸ Berké, Özgür. (2021). A straightforward and Lagrangian proof of the Einsteinian equivalence between the mass and the internal energy (i.e. rest energy). 10.21203/rs.3.rs-642837/v1.
- ⁵⁹ Einstein A (1905) Ist die Trägheit eines Körpers von seinem Energieinhalt abhängig? *Ann Phys* 18:639–643
- ⁶⁰ Einstein A (1908) “Über das Relativitätsprinzip und die aus demselben gezogenen Folgerungen” (PDF). *Jahrbuch der Radioaktivität und Elektronik* 4:411–462 *Jahrbuch der Radioaktivität und Elektronik*, 4: 411–462
- ⁶¹ Einstein, A.. (2021). $E=mc^2$, the most urgent problem. 1. 16-17.
- ⁶² Einstein A, Lorentz HA, Minkowski H et al (1952) *The principle of relativity, a collection of original memoirs on the special and general theory of relativity*, 1st edn. Cover Publication, New York
- ⁶³ Vegh, Janos. (2019). *Classic versus modern computing: analogies with classic versus modern physics*.

⁶⁴ Watson JD, Crick FC (1953) Molecular structure of nucleic acids. *Nature* 171(4356): 737–738

⁶⁵ Thubagere, Anupama & Thachuk, Chris & Berleant, Joseph & Johnson, Robert & Ardelean, Diana & Cherry, Kevin & Qian, Lulu. (2017). Compiler-aided systematic construction of large-scale DNA strand displacement circuits using unpurified components. *Nature Communications*. 8. 14373. 10.1038/ncomms14373.

⁶⁶ Konopik, Michael & Korten, Till & Linke, Heiner & Lutz, Eric. (2021). Solving the subset sum problem with a nonideal biological computer.

⁶⁷ Schrodinger E (1944) *What is life*. Cambridge University Press, UK

⁶⁸ Eck, Dingmar & Wright, Cory. (2020). Mechanist idealisation in systems biology. *Synthese*. 1-21. 10.1007/s11229-020-02816-8.

⁶⁹ Wright, Cory & Eck, Dingmar. (2018). Ontic Explanation Is either Ontic or Explanatory, but Not Both. 5. 997–1029. 10.3998/ergo.12405314.0005.038.

⁷⁰ Rooij, Iris & Wareham, Todd. (2008). Parameterized Complexity in Cognitive Modeling: Foundations, Applications and Opportunities. *Comput. J.* 51. 385-404. 10.1093/comjnl/bxm034.

⁷¹ Uchionkura, Shingo. (2020). Primary and lower secondary students' perceptions of representational practices in science learning: focus on drawing and writing. *International Journal of Science Education*. 42. 1-23. 10.1080/09500693.2020.1849854.

⁷² Rosenblatt, Marcus. (2021). From steady states to dynamics : methods for mechanistic modeling of biological systems with applications to interferon alpha signaling.

Kitano, H. (Ed.). (2001). *Foundations of systems biology*. Cambridge, MA: MIT Press.

⁷³ Röttjers, Lisa & Faust, Karoline. (2018). From hairballs to hypotheses - biological insights from microbial networks. *FEMS microbiology reviews*. 42. 10.1093/femsre/fuy030.

⁷⁴ Gábor, Attila & Banga, Julio. (2015). Robust and efficient parameter estimation in dynamic models of biological systems. *BMC Systems Biology*. 9. 10.1186/s12918-015-0219-2.

⁷⁵ Pradeu, Thomas & Lemoine, Maël & Khelifaoui, Mahdi & Gingras, Yves. (2021). *Philosophy in Science: Can philosophers of science permeate through science and produce scientific knowledge?*.

- ⁷⁶ Kastenhofer, Karen & Molyneux-Hodgson, Susan. (2021). Making Sense of Community and Identity in Twenty-First Century Technoscience. 10.1007/978-3-030-61728-8_1.
- ⁷⁷ Mesarović, M. D. (1968). Systems theory and biology – View of a theoretician. In M. D. Mesarovic (Ed.), *Systems theory and biology. Proceedings of the III systems symposium at case institute of technology* (pp. 59–87). New York: Springer.
- ⁷⁸ Green, Sara & Wolkenhauer, Olaf. (2013). Tracing Organizing Principles: Learning from the History of Systems Biology. *History and Philosophy of the Life Sciences*. 35. 553-576.
- ⁷⁹ Fava, Andrea & Raychaudhuri, Soumya & Rao, Deepak. (2021). The Power of Systems Biology. *Rheumatic Disease Clinics of North America*. 10.1016/j.rdc.2021.04.003.
- ⁸⁰ Nayak, Richa & Chakraborty, Rajkumar & Hasija, Yasha. (2021). System biology and synthetic biology. 10.1016/B978-0-12-821972-0.00012-5.
- ⁸¹ Noble, Denis. (2021). Cellular Darwinism: Regulatory networks, stochasticity, and selection in cancer development. *Progress in Biophysics and Molecular Biology*. 10.1016/j.pbiomolbio.2021.06.007.
- ⁸² Olivença, Daniel & Voit, Eberhard & Pinto, Francisco. (2020). ENaC regulation by phospholipids and DGK explained through mathematical modeling. *Scientific Reports*. 10. 10.1038/s41598-020-70630-w.
- ⁸³ Voit, Eberhard. (2019). Perspective: Dimensions of the scientific method. *PLOS Computational Biology*. 15. e1007279. 10.1371/journal.pcbi.1007279.
- ⁸⁴ Ankeny, Rachel & Leonelli, Sabina. (2020). Model Organisms. 10.1017/9781108593014.
- ⁸⁵ Bechtel, William. (2017). Explaining features of fine-grained phenomena using abstract analyses of phenomena and mechanisms: two examples from chronobiology. *Synthese*. 10.1007/s11229-017-1469-x.
- ⁸⁶ Crombach, Anton & Jaeger, Johannes. (2021). Life's Attractors Continued: Progress in Understanding Developmental Systems Through Reverse Engineering and In Silico Evolution.
- ⁸⁷ Hooker, Cliff. (2013). On the Import of Constraints in Complex Dynamical Systems. *Foundations of Science*. 18. 10.1007/s10699-012-9304-9.

- ⁸⁸ Tee, Sim-Hui. (2019). Computational Construction of the Reality: Abstraction and Exploration-Driven Strategies in Constructing Protein–Protein Interfaces. *Axiomathes*. 29. 10.1007/s10516-018-9391-5.
- ⁸⁹ Raman, Karthik. (2021). An Introduction to Computational Systems Biology: Systems-Level Modelling of Cellular Networks. 10.1201/9780429486951.
- ⁹⁰ Kitano, H. (2002). Computational systems biology. *Nature*, 420, 206–210.
- ⁹¹ Cho, Kwang-Hyun & An, Sugyun & Kang, Junsoo. (2021). Systems biology for reverse aging. *Aging*. 13. 10.18632/aging.203188.
- ⁹² Gross, Fridolin & Green, Sara. (2017). The Sum of the Parts: Large-Scale Modeling in Systems Biology. *Philosophy and Theory in Biology*. 9. 10.3998/ptb.6959004.0009.010.
- ⁹³ Passini, Elisa & Zhou, Xin & Trovato, Cristian & Britton, Oliver & Bueno-Orovio, Alfonso & Rodriguez, Blanca. (2020). The virtual assay software for human in silico drug trials to augment drug cardiac testing. *Journal of Computational Science*. 52. 101202. 10.1016/j.jocs.2020.101202.
- ⁹⁴ Kolodkin, Alexey. (2017). Systems Biology Through the Concept of Emergence. 10.1007/978-3-319-47000-9_17.
- ⁹⁵ Chaiboonchoe, Amphun & Jurkowski, Wiktor & Pellet, J & Glaab, Enrico & Kolodkin, Alexey & Raussel, Antonio & Le Béchech, Antony & Meyniel-Schicklin, Laurène & Ballereau, Stéphane & Crespo, Isaac. (2013). Network analysis for systems biology. Springer book in Systems Biology, Vol. 1: Systems Biology:, Integrative Biology and Simulation Tools.
- ⁹⁶ Oftedal, Gry. (2014). The role of philosophy of science in Responsible Research and Innovation (RRI): the case of nanomedicine. *Life sciences, society and policy*. 10. 5. 10.1186/s40504-014-0005-8.
- ⁹⁷ Istrail, Sorin. (2019). Eric Davidson's Regulatory Genome for Computer Science: Causality, Logic, and Proof Principles of the Genomic cis -Regulatory Code. *Journal of Computational Biology*. 26. 653-684. 10.1089/cmb.2019.0144.
- ⁹⁸ Green, Sara. (2017). Introduction to Philosophy of Systems Biology. 10.1007/978-3-319-47000-9_1.
- ⁹⁹ Rondeau, Elena & Larmonier, Nicolas & Pradeu, Thomas & Bikfalvi, Andreas. (2019). Philosophy of Biology: Characterizing causality in cancer. 8.
- ¹⁰⁰ Ellis, George. (2020). The Causal Closure of Physics in Real World Contexts. *Foundations of Physics*. 50. 10.1007/s10701-020-00366-0.

- ¹⁰¹ Hofmeyr, Jan-Hendrik. (2021). The biochemical factory that autonomously fabricates itself: A systems-biological view of the living cell.
- ¹⁰² Pocheville, Arnaud & Montévil, Maël. (2017). The Hitchhiker's Guide to the Cancer Galaxy. How two critics missed their destination. *Organisms. Journal of Biological Sciences*. 1. 37-48. 10.13133/2532-5876_2.9.
- ¹⁰³ Matthiessen, Dana. (2015). Mechanistic Explanation in Systems Biology: Cellular Networks. *The British Journal for the Philosophy of Science*. 68. 10.1093/bjps/axv011.
- ¹⁰⁴ Mendes, Pedro & Caldwell, Jessica & Camacho, Diogo & de la Fuente, Alberto & Hoops, Stefan & Kamal, Aejaaz & Lee, Christine & Li, Xing & Martins, Ana & Mehrotra, Bharat & Sha, Wei & Shibru, Helen. (2021). *Developing Strategies for Systems Biology*.
- ¹⁰⁵ Hucka, Michael & Finney, A & Sauro, Herbert & Bolouri, H & Doyle, John & Kitano, Hiroaki & Arkin, Adam & Bornstein, B & Bray, D & Cornish-Bowden, Athel & Cuellar, A & Dronov, S & Gilles, E & Ginkel, M & Gor, V & Goryanin, Igor & Hedley, W & Hodgman, T & Hofmeyr, Jan-Hendrik & Wang, J. (2003). The systems biology markup language (SBML): a medium for representation and exchange of biochemical network models. *Bioinformatics*. 19. 524-531.
- ¹⁰⁶ Ali, Md & Parisutham, Vinuselvi & Choubey, Sandeep & Brewster, Robert. (2020). Inherent regulatory asymmetry emanating from network architecture in a prevalent autoregulatory motif. 10.7554/eLife.56517.sa2.
- ¹⁰⁷ Voit, Eberhard. (2013). *Biochemical Systems Theory: A Review*. ISRN Biomathematics. 2013. 10.1155/2013/897658.
- ¹⁰⁸ Sforzini, Susanna & Oliveri, Caterina & Barranger, Audrey & Jha, Awadhesh & Banni, Mohamed & Moore, Michael & Viarengo, Aldo. (2019). Effects of fullerene C60 in blue mussels: Role of mTOR in autophagy related cellular/tissue alterations. *Chemosphere*. 246. 125707. 10.1016/j.chemosphere.2019.125707.
- ¹⁰⁹ Green, Sara & Batterman, Robert. (2021). Making sense of top-down causation: Universality and functional equivalence in physics and biology.
- ¹¹⁰ Auffray, Charles & Noble, Denis & Nottale, Laurent & Turner, Philip. (2020). Progress in integrative systems biology, physiology and medicine: towards a scale-relative biology. *The European Physical Journal A*. 56. 10.1140/epja/s10050-020-00090-3.
- ¹¹¹ Passini, Elisa & Zhou, Xin & Trovato, Cristian & Britton, Oliver & Bueno-Orovio, Alfonso & Rodriguez, Blanca. (2020). The virtual assay software for human in silico drug trials to augment drug cardiac testing. *Journal of Computational Science*. 52. 101202. 10.1016/j.jocs.2020.101202.

- ¹¹² Musuamba, Flora & Rusten, Ine & Lesage, Raphaëlle & Russo, Giulia & Bursi, Roberta & Emili, Luca & Wangorsch, Gaby & Manolis, Efthymios & Karlsson, Kristin & Kulesza, Alexander & Courcelles, Eulalie & Boissel, Jean-Pierre & Rousseau, Cécile & Voisin, Emmanuelle & Alessandrello, Rossana & Curado, Nuno & Dall'ara, Enrico & Rodriguez, Blanca & Pappalardo, Francesco & Geris, Liesbet. (2021). Scientific and regulatory evaluation of mechanistic in silico drug and disease models in drug development: building model credibility. *CPT: Pharmacometrics & Systems Pharmacology*. 10.1002/psp4.12669.
- ¹¹³ Green, Sara & Jones, Nicholaos. (2016). Constraint-Based Reasoning for Search and Explanation: Strategies for Understanding Variation and Patterns in Biology: Constraint-Based Reasoning for Search and Explanation. *Dialectica*. 70. 343-374. 10.1111/1746-8361.12145.
- ¹¹⁴ Vanhoefer, Jakob & Matos, Marta & Pathirana, Dilan & Schälte, Yannik & Hasenauer, Jan. (2021). yaml2sbml: Human-readable and -writable specification of ODE models and their conversion to SBML. *Journal of Open Source Software*. 6. 3215. 10.21105/joss.03215.
- ¹¹⁵ Kolodkin, Alexey & Westerhoff, H.V.. (2011). Parsimony for systems biology: shaving Occam's razor away. *Eur. Commun. Math. Theor. Biol.*. 14. 149-152.
- ¹¹⁶ Hofmeyr, Jan-Hendrik. (2017). Basic Biological Anticipation. 10.1007/978-3-319-31737-3_51-1.
- ¹¹⁷ Carusi, Annamaria. (2017). Enactments of Systems Biology. 10.1007/978-3-319-47000-9_5.
- ¹¹⁸ Tee, Sim-Hui. (2020). Abstraction as an Autonomous Process in Scientific Modeling. *Philosophia*. 48. 10.1007/s11406-019-00092-6.
- ¹¹⁹ Dammann, Olaf. (2020). Etiological Explanations: Illness Causation Theory. 10.1201/9780429184147.
- ¹²⁰ Boon, Mieke. (2017). An engineering paradigm in the biomedical sciences: Knowledge as epistemic tool. *Progress in biophysics and molecular biology*. 129. 10.1016/j.pbiomolbio.2017.04.001.
- ¹²¹ Green, Sara & Batterman, Robert. (2021). Making sense of top-down causation: Universality and functional equivalence in physics and biology.
- ¹²² Love, Bradley. (2021). Levels of biological plausibility. *Philosophical Transactions of the Royal Society B: Biological Sciences*. 376. 20190632. 10.1098/rstb.2019.0632.

- ¹²³ Zednik, Carlos. (2018). Models and Mechanisms in Network Neuroscience. *Philosophical Psychology*. 32. 10.1080/09515089.2018.1512090.
- ¹²⁴ Watts, Fraser & Reiss, Michael. (2017). Holistic Biology: What It Is And Why It Matters. *Zygon(r)*. 52. 419-441. 10.1111/zygo.12339.
- ¹²⁵ Gebharter, Alexander & Koch, Daniel. (2018). Combining Causal Bayes Nets and Cellular Automata: A Hybrid Modelling Approach to Mechanisms. *The British Journal for the Philosophy of Science*. forthcoming. 10.1093/bjps/axy049.
- ¹²⁶ Green, Sara. (2015). Can biological complexity be reverse engineered?. *Studies in history and philosophy of biological and biomedical sciences*. 53. 10.1016/j.shpsc.2015.03.008.
- ¹²⁷ Zhang, Xin. (2021). Are topological explanations really free of mechanisms?. *Theory in Biosciences*. 140. 10.1007/s12064-020-00336-0.
- ¹²⁸ Griffiths, Paul. (2008). In What Sense Does 'Nothing Make Sense Except in the Light of Evolution'?. *Acta biotheoretica*. 57. 11-32. 10.1007/s10441-008-9054-9.
- ¹²⁹ Vaidyanathan, Seetharaman & Harrigan, George & Goodacre, Royston. (2005). *Metabolome Analyses: Strategies for Systems Biology*. 10.1007/b106967.
- ¹³⁰ Sengupta, Pratim & Dickes, Amanda & Farris, Amy. (2018). Toward a Phenomenology of Computational Thinking in STEM Education. 10.1007/978-3-319-93566-9_4.
- ¹³¹ Svensson, Erik. (2017). On reciprocal causation in the evolutionary process. 10.1101/122457.
- ¹³² Griffiths, Paul. (2008). In What Sense Does 'Nothing Make Sense Except in the Light of Evolution'?. *Acta biotheoretica*. 57. 11-32. 10.1007/s10441-008-9054-9.
- ¹³³ Fasan, Rudi & Kan, Jennifer & Zhao, Huimin. (2019). A Continuing Career in Biocatalysis: Frances H. Arnold. *ACS Catalysis*. 9. 10.1021/acscatal.9b02737.
- ¹³⁴ Nath, Bimalendu B.. (2012). Dobzhansky and evolutionary cytogenetics. *Resonance*. 5. 61-65. 10.1007/BF02836842.
- ¹³⁵ Jaeger, J., & Monk, N. (2014). Bioattractors: Dynamical systems theory and the evolution of regulatory processes. *Journal of Physiology*, 592, 2267–2281.
- ¹³⁶ Kurismaa, Andres. (2015). On anticipation as an evolutionary framework: Functional systems perspective. *International Journal of General Systems*. 44. 10.1080/03081079.2015.1032530.

- ¹³⁷ Pigliucci, Massimo. (2008). Do We need an extended evolutionary synthesis?. *Evolution; international journal of organic evolution*. 61. 2743-9. 10.1111/j.1558-5646.2007.00246.x.
- ¹³⁸ Goodwin, Brian. (2020). How the Leopard Changed Its Spots: The Evolution of Complexity. 10.2307/j.ctv14163sj.
- ¹³⁹ Boojari, Mohammad. (2020). Investigating the Evolution and Development of Biological Systems from the Perspective of Thermo-Kinetics and Systems Theory. *Origins of Life and Evolution of Biospheres*. 50. 1-23. 10.1007/s11084-020-09601-0.
- ¹⁴⁰ Tweney, Ryan. (2011). Nancy J. Nersessian: Creating Scientific Concepts. *Science & Education*. 21. 591-596. 10.1007/s11191-011-9361-4.
- ¹⁴¹ Eck, Dingmar & Mennes, Julie. (2016). Design Explanation and Idealization. *Erkenntnis*. 81. 10.1007/s10670-015-9782-6.
- ¹⁴² Collins, H. M., & Evans, R. (2002). The third wave of science studies: Studies of expertise and experience. *Social Studies of Science*, 32, 235–296.
- ¹⁴³ Ross, Wendy & Smith, Sam & Vistic, J. (2020). Collaborative Creativity. 10.1007/978-3-319-98390-5_85-1.
- ¹⁴⁴ Zhang, Xueyan & Wang, Xiaohong & Zhao, Wei. (2020). Social capital and knowledge integration in interdisciplinary research teams: a multilevel analysis. *Management Decision*. ahead-of-print. 10.1108/MD-12-2019-1684.
- ¹⁴⁵ Gross, F. (2015). The relevance of irrelevance: Explanation in systems biology. In P.-A. Braillard & C. Malaterre (Eds.), *Explanation in biology. An enquiry into the diversity of explanatory patterns in the life sciences* (pp. 175–198). Dordrecht: Springer.
- ¹⁴⁶ Leonelli, Sabina. (2014). What Difference Does Quantity Make? On the Epistemology of Big Data in Biology. *Big data & society*. 1. 10.1177/2053951714534395.
- ¹⁴⁷ Vogt, Henrik & Green, Sara. (2020). Personalised Medicine: Problems of Translation into the Human Domain. 10.1007/978-981-15-7728-4_2.
- ¹⁴⁸ Kumari, Dr. (2020). P4 Medicine: A Futuristic Approach to Health and Disease. *International Journal for Research in Applied Science and Engineering Technology*. 8. 679-684. 10.22214/ijraset.2020.31558.
- ¹⁴⁹ Rode, B.M. & Schwendinger, M.G.. (2021). *Orig. Life Evol. Biosph.*
- ¹⁵⁰ Green, Sara. (2017). Introduction to Philosophy of Systems Biology. 10.1007/978-3-319-47000-9_1.

- ¹⁵¹ Kell, Douglas. (2007). Systems Biology, Metabolic Modelling and Metabolomics in Drug Discovery and Development. *Drug discovery today*. 11. 1085-92. 10.1016/j.drudis.2006.10.004.
- ¹⁵² Hill, Camilla & Czauderna, Tobias & Klapperstück, Matthias & Roessner, Ute & Schreiber, Falk. (2015). Metabolomics, Standards, and Metabolic Modeling for Synthetic Biology in Plants. *Frontiers in Bioengineering and Biotechnology*. 3. 167. 10.3389/fbioe.2015.00167.
- ¹⁵³ Butcher, Eugene & Berg, Ellen & Kunkel, Eric. (2004). Systems biology in drug discovery. *Nature biotechnology*. 22. 1253-9. 10.1038/nbt1017.
- ¹⁵⁴ Gross, Fridolin & Green, Sara. (2017). The Sum of the Parts: Large-Scale Modeling in Systems Biology. *Philosophy and Theory in Biology*. 9. 10.3998/ptb.6959004.0009.010.
- ¹⁵⁵ Bunnik, Evelien & Le Roch, Karine. (2013). An Introduction to Functional Genomics and Systems Biology. *Advances in wound care*. 2. 490-498. 10.1089/wound.2012.0379.
- ¹⁵⁶ Peterson, Randall. (2008). Chemical biology and limits of reductionism. *Nature chemical biology*. 4. 635-8. 10.1038/nchembio1108-635.
- ¹⁵⁷ Auffray, Charles & Noble, Denis & Nottale, Laurent & Turner, Philip. (2020). Progress in integrative systems biology, physiology and medicine: towards a scale-relative biology. *The European Physical Journal A*. 56. 10.1140/epja/s10050-020-00090-3.
- ¹⁵⁸ Przedborski, Michelle & Sharon, David & Chan, Steven & Kohandel, Mohammad. (2021). A mean-field approach for modeling the propagation of perturbations in biochemical reaction networks. 10.1101/2021.01.26.428329.
- ¹⁵⁹ Masso, J.R.. (2006). From functional genomics to systems biology. 23. 144-146.
- ¹⁶⁰ From OMICs to systems biology
The field of 'omics' currently polarizes the community of biologists. October 2004 https://www.nature.com/documents/nrg_posters_omics.pdf
- ¹⁶¹ Hans. (2011). Modeling Approaches in Systems Biology, Including Silicon Cell Models. 10.1002/9780470963012.ch2.
- ¹⁶² Noble, Denis. (2020). The role of stochasticity in biological communication processes. *Progress in Biophysics and Molecular Biology*. 10.1016/j.pbiomolbio.2020.09.008.
- ¹⁶³ Sarwar, Dewan & Kalbasi, Reza & Gennari, John & Carlson, Brian & Neal, Maxwell & de Bono, Bernard & Atalag, Koray & Hunter, Peter & Nickerson, David. (2019). Model

annotation and discovery with the Physiome Model Repository. *BMC Bioinformatics*. 20. 10.1186/s12859-019-2987-y.

¹⁶⁴ Saqi, Mansoor & Pellet, Johann & Balaur (Roznovat), Irina & Mazein, Alexander & Ballereau, Stéphane & De Meulder, Bertrand & Auffray, Charles. (2015). *Systems Medicine: The Future of Medical Genomics, Healthcare, and Wellness*. *Methods in molecular biology* (Clifton, N.J.). 1386. 43-60. 10.1007/978-1-4939-3283-2_3.

¹⁶⁵ Estrada-Rivadeneira, Diego. (2017). Sanger sequencing. *The FEBS Journal*. 284. 10.1111/febs.14319.

¹⁶⁶ Mehta, Avni & Hasija, Yasha. (2018). *Bioinformatics Approaches for Genomics and Post Genomics Applications of Anticancer Plants*. 10.1007/978-981-10-8417-1_12.

¹⁶⁷ Voshall, Adam & Moriyama, Etsuko. (2020). Next-generation transcriptome assembly and analysis: Impact of ploidy. *Methods* (San Diego, Calif.). 176. 14-24. 10.1016/j.jymeth.2019.06.001.

¹⁶⁸ Speir, Matthew & Zweig, Ann & Rosenbloom, Kate & Raney, Brian & Paten, Benedict & Nejad, Parisa & Lee, Brian & Learned, Katrina & Karolchik, Donna & Hinrichs, Angie & Heitner, Steve & Harte, Rachel & Haeussler, Maximilian & Guruvadoo, Luvina & Fujita, Pauline & Eisenhart, Christopher & Diekhans, Mark & Clawson, Hiram & Casper, Jonathan & Kent, W.. (2016). The UCSC Genome Browser database: 2016 update. *Nucleic Acids Research*. 44. D717-D725. 10.1093/nar/gkv1275.

¹⁶⁹ Chuna, Kim & Sung, Sanghyun & Kim, Jong-Seo & Lee, Hyunji & Jung, Yoonseok & Shin, Sanghee & Kim, Eunkyong & Seo, Jenny & Kim, Jun & Kim, Daeun & Niida, Hiroyuki & Kim, V. & Park, Daechan & Lee, Junho. (2021). Telomeres reforged with non-telomeric sequences in mouse embryonic stem cells. *Nature Communications*. 12. 10.1038/s41467-021-21341-x.

¹⁷⁰ Xu, Ting & Yang, Zhi & Jiang, Lili & Xing, Xiu-Xia & Zuo, Xi-Nian. (2015). A Connectome Computation System for discovery science of brain. *Science Bulletin*. 60. 10.1007/s11434-014-0698-3.

¹⁷¹ Mamidi, Tarun Karthik Kumar & Wu, Jiande & Hicks, Chindo. (2019). Interactions between Germline and Somatic Mutated Genes in Aggressive Prostate Cancer. *Prostate Cancer*. 2019. 1-11. 10.1155/2019/4047680.

¹⁷² Khansefid, Majid & Pryce, Jennie & Bolormaa, Sunduimijid & Chen, Yizhou & Millen, Catriona & Chamberlain, Amanda & Vander Jagt, Christy & Goddard, Michael. (2018). Comparing allele specific expression and local expression quantitative trait loci and the

influence of gene expression on complex trait variation in cattle. *BMC Genomics*. 19. 10.1186/s12864-018-5181-0.

¹⁷³ Thomas, Torsten & Jack, Gilbert & Meyer, Folker. (2014). *Metagenomics: A Guide From Sampling to Data Analysis*. 10.1201/b16568-16.

¹⁷⁴ Tonkovic, Petar & Kalajdziski, Slobodan & Zdravevski, Eftim & Lameski, Petre & Corizzo, Roberto & Pires, Ivan & Garcia, Nuno & Loncar-Turukalo, Tatjana & Trajkovic, Vladimir. (2020). Literature on Applied Machine Learning in Metagenomic Classification: A Scoping Review. *Biology*. 2020. 453. 10.3390/biology9120453.

¹⁷⁵ Kant, Ravi & Kumar, Abhishek & Sironen, Tarja. (2020). From Microbial Genomics to Metagenomics. *International Journal of Genomics*. 2020. 1-2. 10.1155/2020/9357450.

¹⁷⁶ Dal Molin, Alessandra & Baruzzo, Giacomo & Camillo, Barbara. (2017). Single-Cell RNA-Sequencing: Assessment of Differential Expression Analysis Methods. *Frontiers in Genetics*. 8. 62. 10.3389/fgene.2017.00062.

¹⁷⁷ Farah, Ibrahim & Lalli, Giada & Baker, Darrol & Schumacher, Axel. (2020). A global omics data sharing and analytics marketplace: Case study of a rapid data COVID-19 pandemic response platform. 10.1101/2020.09.28.20203257.

¹⁷⁸ Sayers, Eric & Beck, Jeffrey & Bolton, Evan & Bourexis, Devon & Brister, J. & Canese, Kathi & Comeau, Donald & Funk, Kathryn & Kim, Sunghwan & Klimke, William & Marchler-Bauer, Aron & Landrum, Melissa & Lathrop, Stacy & Lu, Zhiyong & Madden, Thomas & O'Leary, Nuala & Phan, Lon & Rangwala, Sanjida & Schneider, Valerie & Sherry, Stephen. (2020). Database resources of the National Center for Biotechnology Information. *Nucleic Acids Research*. 49. 10.1093/nar/gkaa892.

¹⁷⁹ Banerjee, Arko & Pujari, Arun K & Panigrahi, Chhabi & Pati, Bibudhendu & Nayak, Suvendu & Weng, Tien-Hsiung. (2021). A new method for weighted ensemble clustering and coupled ensemble selection. *Connection Science*. 1-22. 10.1080/09540091.2020.1866496.

¹⁸⁰ Navarro Gonzalez, Jairo & Zweig, Ann & Speir, Matthew & Schmelter, Daniel & Rosenbloom, Kate & Raney, Brian & Powell, Conner & Nassar, Luis & Maulding, Nathan D & Lee, Christopher & Lee, Brian & Hinrichs, Angie & Fyfe, Alastair C & Fernandes, Jason D & Diekhans, Mark & Clawson, Hiram & Casper, Jonathan & Benet-Pagès, Anna & Barber, Galt & Kent, W James. (2020). The UCSC Genome Browser database: 2021 update. *Nucleic Acids Research*. 49. 10.1093/nar/gkaa1070.

¹⁸¹ Picariello, Gianluca & Mamone, Gianfranco & Addeo, Francesco & Ferranti, Pasquale. (2012). Novel Mass Spectrometry-Based Applications of the 'Omic' Sciences

in Food Technology and Biotechnology. Food Technology and Biotechnology. 50. 286-305.

¹⁸² Romero, Pedro. (2004). Bioinformatics: Sequence and Genome Analysis. Briefings in Bioinformatics. 5. 393-396. 10.1093/bib/5.4.393-a.

¹⁸³ Consortium, The. (2000). Gene ontology: Tool for the unification of biology. Nat. Genet.. 25. 25-29.

¹⁸⁴ Stoeckert, Christian & Parkinson, Helen. (2003). The MGED Ontology: A Framework for Describing Functional Genomics Experiments. Comparative and functional genomics. 4. 127-32. 10.1002/cfg.234.

¹⁸⁵ Perez-Riverol, Yasset & Csordas, Attila & Bai, Jingwen & Bernal-Llinares, Manuel & Hewapathirana, Suresh & Kundu, Deepti & Inuganti, Avinash & Griss, Johannes & Mayer, Gerhard & Eisenacher, Martin & Pérez, Enrique & Uszkoreit, Julian & Pfeuffer, Julianus & Sachsenberg, Timo & Yilmaz, Şule & Tiwary, Shivani & Cox, Juergen & Audain, Enrique & Walzer, Mathias & Vizcaino, Juan. (2018). The PRIDE database and related tools and resources in 2019: improving support for quantification data. Nucleic acids research. 47. 10.1093/nar/gky1106.

¹⁸⁶ Zaslavsky, Elena. (2018). Interpretation of an individual functional genomics experiment guided by massive public data. Nature Methods. 15. 10.1038/s41592-018-0218-5.

¹⁸⁷ Doyle, John & Nielsen, Poul. (2021). Systems Biology Markup Language Standard.

¹⁸⁸ Jansson, Andreas & Jirstrand, Mats. (2010). Biochemical modeling with Systems Biology Graphical Notation. Drug discovery today. 15. 365-70. 10.1016/j.drudis.2010.02.012.

¹⁸⁹ Perez-Llamas, Christian & Gundem, Gunes & López-Bigas, Nuria. (2011). Integrative Cancer Genomics (IntOGen) in Biomart. Database : the journal of biological databases and curation. 2011. bar039. 10.1093/database/bar039.

¹⁹⁰ Bierkens, Mariska & van der Linden, Wim & van Bochove, Kees & Weistra, Ward & Fijneman, Remond & Azevedo, Rita & Boiten, Jan-Willem & Beliën, Jeroen & Meijer, Gerrit. (2015). tranSMART. Journal of Clinical Bioinformatics. 5. S9. 10.1186/2043-9113-5-S1-S9.

¹⁹¹ Wang, Caihong & Zhao, Feng-Qi & Liu, Jian-xin & Liu, Hongyun. (2020). Short communication: The essential role of N-glycosylation in the transport activity of bovine peptide transporter 2. Journal of Dairy Science. 103. 10.3168/jds.2019-16858.

- ¹⁹² Yenes, Ronni & Gutrianto, Ryan. (2021). Basic Skills of Men Garuda Basketball Athlete. 10.2991/ahsr.k.210130.007.
- ¹⁹³ Potera, C. (2005). NexBio takes novel approach to influenza prevention by disrupting receptors in airways. *Genetic Engineering News*. 25. 22-22.
- ¹⁹⁴ Rueping, Magnus & Karan, Ram & Sysoev, Maxim & Grötzinger, Stefan & Renn, Dominik & Eppinger, Jörg. (2021). Bioprospecting of Novel Extremozymes From Prokaryotes-The Advent of Culture-Independent Methods. *Frontiers in Microbiology*. 12. 10.3389/fmicb.2021.630013.
- ¹⁹⁵ Gupta, Romi & Nagarajan, Arvindhan & Wajapeyee, Narendra. (2010). Advances in genome-wide DNA methylation analysis. *BioTechniques*. 49. iii-xi. 10.2144/000113493.
- ¹⁹⁶ Paul, DS & Beck, S. (2014). Advances in epigenome-wide association studies for common diseases.. 20.
- ¹⁹⁷ Birney, Ewan & Stamatoyannopoulos, John & Dutta, Anindya & Guigó, Roderic & Gingeras, Thomas & Margulies, Elliott & Deng, Zhiping & Snyder, Michael & Dermitzakis, Emmanouil & Thurman, Robert & Kuehn, Michael & Taylor, Christopher & Neph, Shane & Koch, Christoph & Asthana, Saurabh & Malhotra, Ankit & Adzhubey, Ivan & Greenbaum, Jason & Andrews, Robert & de Jong, Pieter. (2007). Identification and analysis of functional elements in 1% of the human genome by the ENCODE pilot project. *Nature*. 447. 799-816. 10.1038/nature05874.
- ¹⁹⁸ Chadwick, Lisa. (2012). The NIH Roadmap Epigenomics Program data resource. *Epigenomics*. 4. 317-24. 10.2217/epi.12.18.
- ¹⁹⁹ Rauscher, Frank. (2006). It Is Time for a Human Epigenome Project. *Cancer research*. 65. 11229. 10.1158/0008-5472.CAN-65-24-ED1.
- ²⁰⁰ Adams, David & Altucci, Lucia & Antonarakis, Stylianos & Ballesteros, Juan & Beck, Stephan & Bird, Adrian & Bock, Christoph & Boehm, Bernhard & Campo, Elias & Caricasole, Andrea & Dahl, Fredrik & Dermitzakis, Emmanouil & Enver, Tariq & Esteller, Manel & Estivill, Xavier & Ferguson-Smith, Anne & Fitzgibbon, Jude & Flicek, Paul & Schacht, Claudia & Willcocks, Spike. (2012). BLUEPRINT to decode the epigenetic signature written in blood. *Nature biotechnology*. 30. 224-6. 10.1038/nbt.2153.
- ²⁰¹ Strohmalm, Martin & Hassman, Martin & Kosata, Bedrich & Kodíček, Milan. (2008). mMass data miner: An open source alternative for mass spectrometric data analysis. *Rapid communications in mass spectrometry : RCM*. 22. 905-8. 10.1002/rcm.3444.
- ²⁰² Ladd-Acosta, Christine & Aryee, Martin & Ordway, Jared & Feinberg, Andrew. (2010). Comprehensive High-Throughput Arrays for Relative Methylation (CHARM). *Current*

protocols in human genetics / editorial board, Jonathan L. Haines ... [et al.]. Chapter 20. Unit 20.1.1-19. 10.1002/0471142905.hg2001s65.

²⁰³ Aberg, Karolina & Chan, Robin & Xie, Linying & Shabalin, Andrey & Oord, Edwin. (2018). Methyl-CpG-binding domain sequencing: MBD-seq. 10.1007/978-1-4939-7481-8_10.

²⁰⁴ Moran, Sebastian & Arribas, Carles & Esteller, Manel. (2015). Validation of a DNA methylation microarray for 850,000 CpG sites of the human genome enriched in enhancer sequences. *Epigenomics*. 8. 10.2217/epi.15.114.

²⁰⁵ Urban, Lara & Holzer, Andre & Baronas, J. Jotautas & Hall, Michael & Braeuninger, Philipp & Scherm, Michael & Kunz, Daniel & Perera, Surangi & Martín Herranz, Daniel Elías & Tipper, Edward & Salter, Susannah & Stammnitz, Maximilian. (2021). Freshwater monitoring by nanopore sequencing. *eLife Sciences*. 10. 1-27. 10.7554/eLife.61504.

²⁰⁶ Nizamuddin, Sheikh & Koidl, Stefanie & Bhuiyan, Tanja & Werner, Tamara & Biniossek, Martin & Bonvin, Alexandre & Lassmann, Silke & Timmers, HThMarc. (2021). Integrating quantitative proteomics with accurate genome profiling of transcription factors by greenCUT&RUN. *Nucleic acids research*. 10.1093/nar/gkab038.

²⁰⁷ Serre, David & Lee, Byron & Ting, Angela. (2009). MBD-isolated Genome Sequencing provides a high-throughput and comprehensive survey of DNA methylation in the human genome. *Nucleic acids research*. 38. 391-9. 10.1093/nar/gkp992.

²⁰⁸ Ji, Hongkai & Vokes, Steven & Wong, Wing. (2006). A comparative analysis of genome-wide chromatin immunoprecipitation data for mammalian transcription factors. *Nucleic acids research*. 34. e146. 10.1093/nar/gkl803.

²⁰⁹ Assmann, Vera. (2020). CD83: insights into its function and transcriptional regulation in human Tregs.

²¹⁰ Lee, En-Shiun & Sze-To, Antonio & Wong, Andrew & Stashuk, Daniel. (2016). Unsupervised Pattern Discovery in Biosequences Using Aligned Pattern Clustering. *SM Journal of Bioinformatics and Proteomics*. 1. 1008.

²¹¹ Rosa, Stefanie & Shaw, Peter. (2013). Insights into Chromatin Structure and Dynamics in Plants. *Biology*. 2. 1378-410. 10.3390/biology2041378.

²¹² Liu, Fang-Wei & Liao, H.-F & Lin, S.-P & Lu, Yen-Wen. (2018). DNA methylation assay using droplet-based DNA melting curve analysis. *Lab on a Chip*. 18. 10.1039/C7LC01240C.

- ²¹³ Li, Shizhao & Tollefsbol, Trygve. (2020). DNA methylation methods: Global DNA methylation and methylomic analyses. *Methods*. 10.1016/j.ymeth.2020.10.002.
- ²¹⁴ Potter, Dustin & Yan, Pearly & Huang, Tim & Lin, Shili. (2008). Probe signal correction for differential methylation hybridization experiments. *BMC bioinformatics*. 9. 453. 10.1186/1471-2105-9-453.
- ²¹⁵ Decock, Anneleen & Ongenaert, Maté & Cannoodt, Robrecht & Verniers, Kimberly & De Wilde, Bram & Laureys, Geneviève & Roy, Nadine & Berbegall, Ana & Bienertova-Vasku, Julie & Bown, Nick & Clement, Nathalie & Combaret, Valérie & Haber, Michelle & Hoyoux, Claire & Murray, Jayne & Noguera, Rosa & Pierron, Gaele & Schleiermacher, Gudrun & Schulte, Johannes & Vandesompele, Jo. (2015). Methyl-CpG-binding domain sequencing reveals a prognostic methylation signature in neuroblastoma. *Oncotarget*. 7. 10.18632/oncotarget.6477.
- ²¹⁶ Simon, Jeremy & Giresi, Paul & Davis, Ian & Lieb, Jason. (2012). Addendum: Using formaldehyde-assisted isolation of regulatory elements (FAIRE) to isolate active regulatory DNA. *Nature protocols*. 7. 256-67. 10.1038/nprot.2011.444.
- ²¹⁷ Henikoff, Steven & Henikoff, Jorja & Kaya-Okur, Hatice & Ahmad, Kami. (2020). Efficient chromatin accessibility mapping in situ by nucleosome-tethered tagmentation. *eLife*. 9. e63274. 10.7554/elife.63274.
- ²¹⁸ Bodega, Beatrice & Lanzuolo, Chiara. (2021). Capturing Chromosome Conformation Methods and Protocols: Methods and Protocols. 10.1007/978-1-0716-0664-3.
- ²¹⁹ Leblanc, Benjamin & Comet, Itys & Bantignies, Frédéric & Cavalli, Giacomo. (2016). Chromosome Conformation Capture on Chip (4C): Data Processing. *Methods in molecular biology (Clifton, N.J.)*. 1480. 243-261. 10.1007/978-1-4939-6380-5_21.
- ²²⁰ Belton, Jon-Matthew & Dekker, Job. (2015). Chromosome Conformation Capture Carbon Copy (5C) in Budding Yeast. *Cold Spring Harbor protocols*. 2015. pdb.prot085191. 10.1101/pdb.prot085191.
- ²²¹ Davis, Sage & Hollin, Thomas & Lenz, Todd & Le Roch, Karine. (2021). Three-dimensional chromatin in infectious disease-A role for gene regulation and pathogenicity?. *PLoS Pathogens*. 17. e1009207. 10.1371/journal.ppat.1009207.
- ²²² Strobino, Maude & Wenda, Joanna & Padayachy, Laura & Steiner, Florian. (2020). Loss of histone H3.3 results in DNA replication defects and altered origin dynamics in *C. elegans*. *Genome Research*. 30. 1740-1751. 10.1101/gr.260794.120.
- ²²³ Gaffney, Daniel & McVicker, Graham & Pai, Athma & Fondufe-Mittendorf, Yvonne & Lewellen, Noah & Michelini, Katelyn & Widom, Jonathan & Gilad, Yoav & Pritchard,

Jonathan. (2012). Controls of Nucleosome Positioning in the Human Genome. *PLoS genetics*. 8. e1003036. 10.1371/journal.pgen.1003036.

²²⁴ Li, Yang & Zhou, Shiguo & Schwartz, David & Ma, Jian. (2016). Allele-Specific Quantification of Structural Variations in Cancer Genomes. *Cell Systems*. 3. 10.1016/j.cels.2016.05.007.

²²⁵ Cusanovich, Darren & Pavlovic, Bryan & Pritchard, Jonathan & Gilad, Yoav. (2014). The Functional Consequences of Variation in Transcription Factor Binding. *PLoS genetics*. 10. e1004226. 10.1371/journal.pgen.1004226.

²²⁶ Jeong, Jae-Sik. (2011). A Finite Mixture Model for Gene Expression and Methylation Profiles in a Bayesian Framework. *Korean Journal of Applied Statistics*. 24. 609-622. 10.5351/KJAS.2011.24.4.609.

²²⁷ Misawa, Kiyoshi & Mochizuki, Daiki & Imai, Atsushi & Mima, Masato & Misawa, Yuki & Kanazawa, Takeharu & Carey, Thomas & Mineta, Hiroyuki. (2016). Prognostic value of aberrant promoter hypermethylation of tumor-related genes in early-stage head and neck cancer. *Oncotarget*. 7. 10.18632/oncotarget.8317.

²²⁸ Roller, Maša & Stamper, Ericca & Villar, Diego & Izuogu, Osagie & Martin, Fergal & Redmond, Aisling & Ramachandran, Raghavendra & Harewood, Louise & Odom, Duncan & Flicek, Paul. (2021). LINE retrotransposons characterize mammalian tissue-specific and evolutionarily dynamic regulatory regions. *Genome Biology*. 22. 10.1186/s13059-021-02260-y.

²²⁹ Aïssi, Dylan & Dennis, Jessica & Ladouceur, Martin & Truong, Vinh & Zwingerman, Nora & Rocañín Arjó, Ares & Germain, Marine & Paton, Tara & Morange, Pierre & Gagnon, France & Trégouët, David-Alexandre. (2014). Genome-Wide Investigation of DNA Methylation Marks Associated with FV Leiden Mutation. *PLoS ONE*. 9. e108087. 10.1371/journal.pone.0108087.

²³⁰ Parkinson, Helen & Sarkans, Ugis & Shojatalab, M & Abeygunawardena, N & Contrino, Sergio & Coulson, R & Farne, Anna & Lara, G & Holloway, E & Kapushesky, Misha & Lilja, P & Mukherjee, G & Özçimen, Ahmet & Rayner, Tim & Rocca-Serra, P & Sharma, A & Sansone, S & Brazma, A. (2005). ArrayExpress - A public repository for microarray gene expression data at the EBI. *Nucleic acids research*. 33. D553-5. 10.1093/nar/gki056.

²³¹ Barrett, T. & Troup, D.B. & Wilhite, S.E. & Ledoux, P. & Evangelista, C. & Kim, I.F.. (2012). NCBI GEO: archive for functional genomics data sets - 10years on. *Nucleic Acids Res*. 39. 10.1093/nar/gks1193.

- ²³² Jr, Christian & Quackenbush, John & Brazma, Alvis & Ball, Catherine. (2004). Minimum information about a functional genomics experiment: The state of microarray standards and their extension to other technologies. *Drug Discovery Today: TARGETS*. 3. 159-164. 10.1016/S1741-8372(04)02435-1.
- ²³³ Guo, Kai & Eid, Stephanie & Elzinga, Sarah & Pacut, Crystal & Feldman, Eva & Hur, Junguk. (2020). Genome-wide profiling of DNA methylation and gene expression identifies candidate genes for human diabetic neuropathy. *Clinical Epigenetics*. 12. 10.1186/s13148-020-00913-6.
- ²³⁴ Mohapatra, Saroj & Krishnan, Arjun. (2011). Microarray Data Analysis. *Methods in molecular biology (Clifton, N.J.)*. 678. 27-43. 10.1007/978-1-60761-682-5_3.
- ²³⁵ Jiang, Chunjie & Ding, Na & Li, Junyi & Jin, Xiyun & Li, Lili & Pan, Tao & Huo, Caiqin & Li, Yongsheng & Xu, Juan & Li, Xia. (2018). Landscape of the long non-coding RNA transcriptome in human heart. *Briefings in Bioinformatics*. 20. 10.1093/bib/bby052.
- ²³⁶ Yu & Yang, Xiaofei & Luo, Haitao. (2013). Long non-coding RNAs: Insights into the biological property. *Chinese Science Bulletin (Chinese Version)*. 58. 2779. 10.1360/972012-1023.
- ²³⁷ Cao, Yunshan & Li, Yahong & Wu, Mianmian & Song, Jiyang & Zhang, Min & Duan, Yichao & Jiang, Kaiyu & Zhou, Xing & Zhang, Yan. (2019). EXPRESS: RNA-sequencing analysis of gene expression in a rat model of acute right heart failure. *Pulmonary Circulation*. 10. 204589401987939. 10.1177/2045894019879396.
- ²³⁸ Mercer, Tim & Gerhardt, Daniel & Dinger, Marcel & Crawford, Joanna & Trapnell, Cole & Jeddellouh, Jeffrey & Mattick, John & Rinn, John. (2011). Targeted RNA sequencing reveals the deep complexity of the human transcriptome. *Nature biotechnology*. 30. 99-104. 10.1038/nbt.2024.
- ²³⁹ Lareau, Caleb & Ludwig, Leif & Muus, Christoph & Gohil, Satyen & Zhao, Tongtong & Chiang, Zachary & Pelka, Karin & Verboon, Jeffrey & Luo, Wendy & Christian, Elena & Rosebrock, Daniel & Getz, Gad & Boland, Genevieve & Chen, Fei & Buenrostro, Jason & Hacohen, Nir & Wu, Catherine & Aryee, Martin & Regev, Aviv & Sankaran, Vijay. (2020). Massively parallel single-cell mitochondrial DNA genotyping and chromatin profiling. *Nature Biotechnology*. 10.1038/s41587-020-0645-6.
- ²⁴⁰ Brazma, Alvis & Hingamp, Pascal & Quackenbush, John & Sherlock, Gavin & Spellman, Paul & Stoeckert, Chris & Aach, John & Ansorge, Wilhelm & Ball, Catherine & Causton, Helen & Gaasterland, Terry & Glenisson, Patrick & Holstege, Frank & Kim, Irene & Markowitz, Victor & Matese, John & Parkinson, Helen & Robinson, Alan &

Sarkans, Ugis & Vingron, Martin. (2001). Minimum information about a microarray experiment (MIAME) - Toward standards for microarray data. *Nature Genetics*. 29. 10.1038/ng1201-365.

²⁴¹ Voshall, Adam & Moriyama, Etsuko. (2020). Next-generation transcriptome assembly and analysis: Impact of ploidy. *Methods (San Diego, Calif.)*. 176. 14-24. 10.1016/j.ymeth.2019.06.001.

²⁴² Gelfond, Jonathan & Ibrahim, Joseph & Chen, Ming-Hui & Sun, Wei & Lewis, Kaitlyn & Kinahan, Sean & Hibbs, Matthew & Buffenstein, Rochelle. (2015). Homology cluster differential expression analysis for interspecies mRNA-Seq experiments. *Statistical Applications in Genetics and Molecular Biology*. 14. 10.1515/sagmb-2014-0056.

²⁴³ K, Gary & Bumgarner, Roger & B, Brian & Triche, Timothy & Dowidar, Naeem & L, Dwayne & P, H & S, Sean & George, Renee & T, Tammy & J, Jeffrey & Maysuria, Malini. (2008). Direct multiplexed measurement of gene expression with color-coded probe pairs (vol 26, pg 317, 2008). *Nature Biotechnology - NAT BIOTECHNOL*. 26. 709-709. 10.1038/nbt0608-709c

²⁴⁴ Ren, Conglin & Li, Mingshuang & Du, Weibin & Lü, Jianlan & Zheng, Yang & Xu, Haipeng & Quan, Renfu. (2020). Comprehensive Bioinformatics Analysis Reveals Hub Genes and Inflammation State of Rheumatoid Arthritis. *BioMed Research International*. 2020. 1-13. 10.1155/2020/6943103.

²⁴⁵ Kang, Seoyoung & Larbi, Daniel & Andrade, Monica & Reardon, Sara & Reh, Thomas & Wohl, Stefanie. (2021). A Comparative Analysis of Reactive Müller Glia Gene Expression After Light Damage and microRNA-Depleted Müller Glia—Focus on microRNAs. *Frontiers in Cell and Developmental Biology*. 8. 620459. 10.3389/fcell.2020.620459.

²⁴⁶ Zhang, Zhicai & Liu, Jianxiang & Fan, Jiaming & Huang, Shifeng & Zhang, Linghuan & Zhang, Bo & Wang, Xi & Feng, Yixiao & Ye, Zhenyu & Zhao, Ling & Cao, Daigui & Yang, Lijuan & Pakvasa, Mikhail & Liu, Bin & Wagstaff, William & Wu, Xiaoxing & Luo, Huaxiu & Zhang, Jing & He, Tong-Chuan. (2019). lncRNA Rmst acts as an important mediator of BMP9-induced osteogenic differentiation of mesenchymal stem cells (MSCs) by antagonizing Notch-targeting microRNAs. *Aging*. 11. 10.18632/aging.102583.

²⁴⁷ Bernstein, David & Sulheim, Snorre & Almaas, Eivind & Segrè, Daniel. (2021). Addressing uncertainty in genome-scale metabolic model reconstruction and analysis. *Genome Biology*. 22. 10.1186/s13059-021-02289-z.

- ²⁴⁸ Wei, Zhengxi & Zhao, Jinghua & Niebler, Jake & Hao, Jian-Jiang & Merrick, Alex & Xia, Menghang. (2020). Quantitative Proteomic Profiling of Mitochondrial Toxicants in a Human Cardiomyocyte Cell Line. *Frontiers in Genetics*. 11. 719. 10.3389/fgene.2020.00719.
- ²⁴⁹ Duarte, Jessica & Goosen, Ryan & Lawry, Peter & Blackburn, Jonathan. (2018). PMA: Protein Microarray Analyser, a user-friendly tool for data processing and normalization. *BMC Research Notes*. 11. 156. 10.1186/s13104-018-3266-0.
- ²⁵⁰ Ronkainen, Niina & Okon, Stanley. (2014). Nanomaterial-Based Electrochemical Immunosensors for Clinically Significant Biomarkers. *Materials*. 7. 4669-4709. 10.3390/ma7064669.
- ²⁵¹ Ezkurdia, Iakes & Juan, David & Rodríguez, Jose Manuel & Frankish, Adam & Deikhans, Mark & Harrow, Jennifer & Vázquez, Jesús & Valencia, Alfonso & Tress, Michael. (2014). The shrinking human protein coding complement: are there fewer than 20,000 genes?. 10.1101/001909.
- ²⁵² Dahabiyeh, Lina. (2018). The Discovery of Protein Biomarkers in Pre-eclampsia: The Promising Role of Mass Spectrometry.
- ²⁵³ Conrad, Tim & Genzel, Martin & Cvetkovic, Nada & Wulkow, Niklas & Leichtle, Alexander & Vybiral, Jan & Kutyniok, Gitta & Schütte, Christof. (2017). Sparse Proteomics Analysis - A Compressed Sensing-based Approach for Feature Selection and Classification of High-Dimensional Proteomics Mass Spectrometry Data. *BMC Bioinformatics*. 18. 160-180. 10.1186/s12859-017-1565-4.
- ²⁵⁴ Li, Xu & Wang, Wenqi & Chen, Junjie. (2015). From pathways to networks: Connecting dots by establishing protein-protein interaction networks in signaling pathways using affinity purification and mass spectrometry. *PROTEOMICS*. 15. 10.1002/pmic.201400147.
- ²⁵⁵ Matthiesen, Rune & Bunkenborg, Jakob. (2019). Introduction to Mass Spectrometry-Based Proteomics. 10.1007/978-1-4939-9744-2_1.
- ²⁵⁶ Tomlinson, Chris & Barton, Geraint & Woodbridge, Mark & Butcher, Sarah. (2013). XperimentR: Painless annotation of a biological experiment for the laboratory scientist. *BMC bioinformatics*. 14. 8. 10.1186/1471-2105-14-8.
- ²⁵⁷ Gotelli, Nicholas & Ellison, Aaron & Ballif, Bryan. (2012). Environmental Proteomics, Biodiversity Statistics, and Food-Web Structure. *Trends in ecology & evolution*. 27. 436-42. 10.1016/j.tree.2012.03.001.

- ²⁵⁸ Garavito, Andrea & González-Muñoz, Andrea & Mosquera, Jeanneth & Alvarez, Astrid & Lopez-Alvarez, Diana & Cristancho, Marco. (2017). Latin American biodiversity and perspectives to study it using 'omics' technologies. *Mexican journal of biotechnology*. 2. 89-129. 10.29267/mxjb.2017.2.2.98.
- ²⁵⁹ Chen, Chen & Hou, Jie & Tanner, John & Cheng, Jianlin. (2020). Bioinformatics Methods for Mass Spectrometry-Based Proteomics Data Analysis. *International Journal of Molecular Sciences*. 21. 2873. 10.3390/ijms21082873.
- ²⁶⁰ Guthals, Adrian & Bandeira, Nuno. (2012). Peptide Identification by Tandem Mass Spectrometry with Alternate Fragmentation Modes. *Molecular & cellular proteomics : MCP*. 11. 550-7. 10.1074/mcp.R112.018556.
- ²⁶¹ Hao, Junheng & Ju, Chelsea & Chen, Muhao & Sun, Yizhou & Zaniolo, Carlo & Wang, Wei. (2020). Bio-JOIE: Joint Representation Learning of Biological Knowledge Bases. 10.1101/2020.06.15.153692.
- ²⁶² Goonewardene, Kalhari & Karu, Naama & Ahmed, Khawaja & Popowich, Shelly & Chow-Lockerbie, Betty & Ayalew, Lisanework & Karunarathna, Ruwani & Gunawardana, Thushari & Liu, Mengying & Tikoo, Suresh & Foldvari, Marianna & Willson, Philip & Mandal, Rupasri & Wishart, David & Gomis, Susantha. (2021). CpG-ODN induced antimicrobial immunity in neonatal chicks involves a substantial shift in serum metabolic profiles. *Scientific Reports*. 11. 10.1038/s41598-021-88386-2.
- ²⁶³ Deutsch, Leon & Osredkar, Damjan & Plavec, Janez & Stres, Blaz. (2021). Spinal Muscular Atrophy after Nusinersen Therapy: Improved Physiology in Pediatric Patients with No Significant Change in Urine, Serum, and Liquor 1H-NMR Metabolomes in Comparison to an Age-Matched, Healthy Cohort. *Metabolites*. 11. 206. 10.3390/metabo11040206.
- ²⁶⁴ Fu, Lihao & Zhang, Jianzhi & Si, Tong. (2020). Recent advances in high-throughput mass spectrometry that accelerates enzyme engineering for biofuel research. *BMC Energy*. 2. 10.1186/s42500-020-0011-8.
- ²⁶⁵ Wixon, Joanne & Kell, Douglas. (2000). Website Review: The Kyoto Encyclopedia of Genes and Genomes—KEGG. <http://www.genome.ad.jp/kegg/>. *Yeast*. 17. 48 - 55. 10.1002/(sici)1097-0061(200004)17:1<48::aid-yea2>3.0.co;2-h.
- ²⁶⁶ Arentz, G. & Mittal, Parul & Zhang, Chao & Ho, Y.-Y & Briggs, Matthew & Winderbaum, Lyron & Hoffmann, Marcus & Hoffmann, Peter. (2017). Applications of Mass Spectrometry Imaging to Cancer. 10.1016/bs.acr.2016.11.002.

- ²⁶⁷ Arioli, Angelica & Dagliati, Arianna & Geary, Bethany & Peek, Niels & Kalra, Philip & Whetton, Anthony & Geifman, Nophar. (2021). OptiMissP: A dashboard to assess missingness in proteomic data-independent acquisition mass spectrometry. PLOS ONE. 16. e0249771. 10.1371/journal.pone.0249771.
- ²⁶⁸ Stettin, Daniel & Poulin, Remington & Pohnert, Georg. (2020). Metabolomics Benefits from Orbitrap GC–MS—Comparison of Low- and High-Resolution GC–MS. *Metabolites*. 10. 143. 10.3390/metabo10040143.
- ²⁶⁹ Dong E, Bouatra S, Mandal R, Sinelnikov I, Xia J, Jia L, Cruz JA, Lim E, Sobsey CA, Shrivastava S, Huang P, Liu P, Fang L, Peng J, Fradette R, Cheng D, Tzur D, Clements M, Lewis A, De Souza A, Zuniga A, Dawe M, Xiong Y, Clive D, Greiner R, Nazyrova A, Shaykhutdinov R, Li L, Vogel HJ, Forsythe I (2009) HMDB: a knowledgebase for the human metabolome. *Nucleic Acids Res* 37(Database issue):D603–D610
- ²⁷⁰ Yu, Zhonghao & Kastenmüller, Gabi & He, Ying & Belcredi, Petra & Möller, Gabriele & Prehn, Cornelia & Mendes, Joaquim & Wahl, Simone & Roemisch-Margl, Werner & Ceglarek, Uta & Polonikov, Alexey & Dahmen, Norbert & Prokisch, Holger & Xie, Lu & Li, Yixue & Wichmann, H. & Peters, Annette & Kronenberg, Florian & Suhre, Karsten & Wang-Sattler, Rui. (2011). Differences between Human Plasma and Serum Metabolite Profiles. *PloS one*. 6. e21230. 10.1371/journal.pone.0021230.
- ²⁷¹ Gomez-Cambronero, Julian & Frohman, Michael. (2020). Lipid Signaling in Human Diseases. 10.1007/978-3-030-33668-4.
- ²⁷² Xu, Peiyu & Huang, Sijie & Zhang, Huibing & Mao, Chunyou & Zhou, X. & Cheng, Xi & Simon, Icaro & Shen, Dan-Dan & Yen, Hsin-Yung & Robinson, Carol & Harpsøe, Kasper & Svensson, Bo & Guo, Jia & Jiang, H. & Gloriam, David & Melcher, Karsten & Jiang, Yi & Zhang, Yan & Xu, Eric. (2021). Structural insights into the lipid and ligand regulation of serotonin receptors. *Nature*. 592. 10.1038/s41586-021-03376-8.
- ²⁷³ Almeida, Idália & Magalhães, S. & Nunes, Alexandra. (2021). Lipids: biomarkers of healthy aging. *Biogerontology*. 22. 10.1007/s10522-021-09921-2.
- ²⁷⁴ Nilsson, Anders & Andersson, Mats & Sjöbom, Ulrika & Hellgren, Gunnel & Lundgren, Pia & Pivodic, Aldina & Hellström, Ann. (2021). Sphingolipidomics of serum in extremely preterm infants: Association between low sphingosine-1-phosphate levels and severe retinopathy of prematurity. *Biochimica et biophysica acta. Molecular and cell biology of lipids*. 1866. 158939. 10.1016/j.bbalip.2021.158939.
- ²⁷⁵ Wishart, David & Knox, Craig & Guo, An Chi & Eisner, Roman & Young, Nelson & Gautam, Bijaya & Hau, David & Psychogios, Nick & Dong, Edison & Bouatra, Souhaila

& Mandal, Rupasri & Sinelnikov, Igor & Xia, Jianguo (Jeff) & Jia, Leslie & Cruz, Joseph & Lim, Emilia & Sobsey, Constance & Shrivastava, Savita & Huang, Paul & Forsythe, Ian. (2008). HMDB: A knowledgebase for the human metabolome. *Nucleic acids research*. 37. D603-10. 10.1093/nar/gkn810.

²⁷⁶ Griffin, Emily & Costa, Kaylie & Aristizabal-Henao, Juan & Napolitano, Michael & Hunter, Margaret & Ferrante, Jason & Bowden, John. (2021). Lipidomics reveals specific lipid molecules associated with cold stress syndrome in the Florida manatee (*Trichechus manatus latirostris*). *Marine Biology*. 168. 10.1007/s00227-021-03879-y.

²⁷⁷ Taylor, Nadine & White, Thomas & Viant, Mark. (2017). Defining the Baseline and Oxidant Perturbed Lipidomic Profiles of *Daphnia magna*. *metabolites*. 7. 10.3390/metabo7010011.

²⁷⁸ Chaurand, Pierre & Cornett, Dale & Angel, Peggi & Caprioli, Richard. (2011). From Whole-body Sections Down to Cellular Level, Multiscale Imaging of Phospholipids by MALDI Mass Spectrometry. *Molecular & cellular proteomics : MCP*. 10. O110.004259. 10.1074/mcp.O110.004259.

²⁷⁹ Li, Lili & Ma, Shuangshuang & Wang, Daijie & Chen, Long & Wang, Xiao. (2019). Plasma metabolomics analysis of endogenous and exogenous metabolites in the rat after administration of *Lonicerae Japonicae Flos*. *Biomedical Chromatography*. 34. 10.1002/bmc.4773.

²⁸⁰ Oliver Machate (2021). Evidence for antifouling biocides as one of the limiting factors for the recovery of macrophyte communities in lakes of Schleswig-Holstein. *Environmental Sciences Europe*. 33. 10.1186/s12302-021-00500-3.

²⁸¹ Hyotylainen, Tuulia & Bondia-Pons, Isabel & Oresic, Matej. (2013). Lipidomics in nutrition and food research. *Molecular nutrition & food research*. 57. 10.1002/mnfr.201200759.

²⁸² Dennis, Edward. (2016). Lipid Cell Signaling, Enzymes, LIPID MAPS, and Mediators of Inflammation. *The Journal of biological chemistry*. 291. 10.1074/jbc.X116.723791.

²⁸³ Hsieh, Wei-Yuan & Williams, Kevin & Su, Baolong & Bensinger, Steven. (2021). Profiling of mouse macrophage lipidome using direct infusion shotgun mass spectrometry. *STAR Protocols*. 2. 100235. 10.1016/j.xpro.2020.100235.

²⁸⁴ Meyer-Base, Anke & Lespinats, Sylvain. (2010). Novel Systems Biology and Computational Methods for Lipidomics. *Proceedings of SPIE - The International Society for Optical Engineering*. 7703. 10.1117/12.849906.

- ²⁸⁵ Gehlenborg, Nils & O' Donoghue, Sean & Baliga, Nitin & Goesmann, Alexander & Hibbs, Matthew & Kohlbacher, Oliver & Neuweger, Heiko & Schneider, Reinhard & Tenenbaum, Dan & Gavin, Anne-Claude. (2010). Visualization of omics data for system biology. *Nature methods*. 7. S56-68. 10.1038/nmeth.1436.
- ²⁸⁶ Ballereau, Stéphane & Glaab, Enrico & Kolodkin, Alexey & Chaiboonchoe, Amphun & Biryukov, M & Vlassis, Nikos & Ahmed, Hassan & Pellet, J & Baliga, Nitin & Hood, L & Schneider, R & Balling, Rudi & Auffray, Charles. (2013). *Systems Biology: Integrative Biology and Simulation Tools*.
- ²⁸⁷ Kepczynska, Malgorzata & Wargent, Edward & O'Dowd, Jacqueline & Hislop, David & Arch, Jonathan & Stocker, Claire. (2021). Diabetes Research and Metabolism Metabolic programming of adipose tissue in female C57Bl/6 mice offspring by maternal high fat feeding and obesity results in greater adaptability to dietary fat content and resistance to high fat diet-induced glucose intolerance compared to male mice. *Diabetes Research and Metabolism*. 10.36879/DRM.21.000106.
- ²⁸⁸ Sutherland, Ben & Prokkola, Jenni & Audet, Céline & Bernatchez, Louis. (2019). Sex-Specific Co-expression Networks and Sex-Biased Gene Expression in the Salmonid Brook Charr *Salvelinus fontinalis*. *G3-Genes Genomes Genetics*. 9. g3.200910.2018. 10.1534/g3.118.200910.
- ²⁸⁹ Henriques, Rui & Ferreira, Francisco & C . Madeira, Sara. (2017). BicPAMS: Software for biological data analysis with pattern-based biclustering. *BMC Bioinformatics*. 18. 10.1186/s12859-017-1493-3.
- ²⁹⁰ Alonso, Roberto & Salavert Torres, Francisco & Garcia-Garcia, F. & Carbonell-Caballero, J. & Bleda, Marta & Garcia-Alonso, Luz. (2015). Babelomics 5.0: Functional interpretation for new generations of genomic data. 43. W117-W121.
- ²⁹¹ Appasani, Krishnarao & Southern, Edwin. (2007). *Bioarrays: From Basics to Diagnostics*. 10.1007/978-1-59745-328-8.
- ²⁹² Antfolk, Daniel & Antila, Christian & Kempainen, Kati & Landor, Sebastian & Sahlgren, Cecilia. (2019). Decoding the PTM-switchboard of Notch. *Biochimica et Biophysica Acta (BBA) - Molecular Cell Research*. 1866. 10.1016/j.bbamcr.2019.07.002.
- ²⁹³ Dowluru, Kaladhar SVGK. (2013). *The Elements of Statistical Learning in Colon Cancer Datasets: Data Mining, Inference and Prediction*.
- ²⁹⁴ Robertson, Claire. (2015). The extracellular matrix in breast cancer predicts prognosis through composition, splicing and crosslinking. *Experimental Cell Research*. 343. 10.1016/j.yexcr.2015.11.009.

- ²⁹⁵ Frøssing, Laurits & Silberbrandt, Alexander & Bülow, Anna & Klein, Ditte & Christensen, Marcus & Backer, Vibeke & Baines, Katherine & Porsbjerg, Celeste. (2021). Airway gene expression identifies subtypes of Type 2 inflammation in severe asthma. *Clinical & Experimental Allergy*. 10.1111/cea.13966
- ²⁹⁶ Park, Heung-Woo & Weiss, Scott. (2020). Understanding the Molecular Mechanisms of Asthma through Transcriptomics. *Allergy, Asthma & Immunology Research*. 12. 399. 10.4168/aair.2020.12.3.399.
- ²⁹⁷ Bellman R (1961) Adaptive control processes. Princeton University Press, Princeton
- ²⁹⁸ Wang, Shaohua & Hu, Qingwu & Wang, Fengzhu & Ai, Mingyao & Zhong, Ruofei. (2017). A Microtopographic Feature Analysis-Based LiDAR Data Processing Approach for the Identification of Chu Tombs. *Remote Sensing*. 9. 880. 10.3390/rs9090880.
- ²⁹⁹ Noviantoro, Tri. (2021). APPLYING DATA MINING TECHNIQUES TO INVESTIGATE ONLINE SHOPPER PURCHASE INTENTION BASED ON CLICKSTREAM DATA.
- ³⁰⁰ Okun, Oleg. (2011). Unsupervised Feature Selection. 10.4018/978-1-60960-557-5.ch014.
- ³⁰¹ Tsamardinos, Ioannis & Aliferis, Constantin. (2002). Towards Principled Feature Selection: Relevancy, Filters and Wrappers. in *Proceedings of the Ninth International Workshop on Artificial Intelligence and Statistics*.
- ³⁰² Michmizos, Konstantinos & Argyropoulos, Christos & Aidinis, Vassilis & Nikiforidis, George & Berberidis, Kostas. (2007). A proposal of a novel segmentation technique of microarray images.
- ³⁰³ A. Badr, Ameer & Karim, Alia. (2021). CatBoost Machine Learning Based Feature Selection for Age and Gender Recognition in Short Speech Utterances. *International Journal of Intelligent Engineering and Systems*. Vol.14. 2021. 10.22266/ijjes2021.0630.14.
- ³⁰⁴ Dincer, Ayse & Janizek, Joseph & Lee, Su-In. (2020). Adversarial deconfounding autoencoder for learning robust gene expression embeddings. *Bioinformatics*. 36. i573-i582. 10.1093/bioinformatics/btaa796.
- ³⁰⁵ Abdulwahhab, Ahmed & Mokhtar, Makhfudzah & Saripan, M Iqbal & Abu Bakar, Muhammad Hafiz. (2015). Integrated framework of feature selection from microarray data for classification. *Journal of Theoretical and Applied Information Technology*. 73. 190-201.

- ³⁰⁶ Katchborian Neto, Albert & Santos, Wanderleya & Nicacio, Karen & Corrêa, José & Murgu, Michael & Martins, Thaís & Gomes, Dawidson & Góes, Alfredo & Soares, Marisi & Dias, Danielle & Chagas--Paula, Daniela Aparecida & Paula, Ana. (2020). Neuroprotective potential of Ayahuasca and untargeted metabolomics analyses: Applicability to Parkinson's disease. *Journal of ethnopharmacology*. 255. 112743. 10.1016/j.jep.2020.112743.
- ³⁰⁷ Li, Tianjun & Chen, Long & Lu, Xiliang. (2020). An Alternating Direction Minimization based denoising method for extracted ion chromatogram. *Chemometrics and Intelligent Laboratory Systems*. 206. 104138. 10.1016/j.chemolab.2020.104138.
- ³⁰⁸ Tian, Xiaobo & Permentier, Hjalmar & Bischoff, Rainer. (2021). Chemical isotope labeling for quantitative proteomics. *Mass Spectrometry Reviews*. 10.1002/mas.21709.
- ³⁰⁹ Shang, Ronghua & Chang, Jiangwei & Jiao, Licheng & Xue, Yu. (2019). Unsupervised feature selection based on self-representation sparse regression and local similarity preserving. *International Journal of Machine Learning and Cybernetics*. 10. 10.1007/s13042-017-0760-y.
- ³¹⁰ Hall, Mark. (2000). *Correlation-Based Feature Selection for Machine Learning*. Department of Computer Science. 19.
- ³¹¹ Gupta, Subodhini & Jivani, Anjali. (2021). Feature Selection through Clustering to Classify High-Dimensional Data. 10.1007/978-981-15-9873-9_13.
- ³¹² Biswal, Chandra & Pani, Dr. Subhendu & Dash, Sujata. (2021). A Comparative Analysis of Classifiers Using Particle Swarm Optimization-Based Feature Selection.
- ³¹³ Rahmawati, Rahmawati. (2015). Kernel Bayesian Based Classification for Microarray data.
- ³¹⁴ Hambali, Moshood & Oladele, Tinuke & S., Adewole. (2020). Microarray cancer feature selection: Review, challenges and research directions. *International Journal of Cognitive Computing in Engineering*. 1. 78-97. 10.1016/j.ijcce.2020.11.001.
- ³¹⁵ Singh, Rachhpal. (2018). A Gene Expression Data Classification and Selection Method using Hybrid Meta-heuristic technique. *ICST Transactions on Scalable Information Systems*. 7. 159917. 10.4108/eai.13-7-2018.159917.
- ³¹⁶ Kadhim, Thair & Hasan, Mohammed & Zainudin, Suhaila & Bakar, Azuraliza. (2019). Combination between DE and SVM to enhance Protein Structure Prediction based on Secondary Structural information. *International Journal of Engineering & Technology*. 8. 478. 10.14419/ijet.v8i4.19619

- ³¹⁷ Bolón-Canedo, Verónica & Sánchez-Marroño, Noelia & Alonso-Betanzos, Amparo. (2012). An ensemble of filters and classifiers for microarray data classification. *Pattern Recognition*. 45. 531-539. 10.1016/j.patcog.2011.06.006.
- ³¹⁸ Felici, Giovanni & Tripathi, Kumar Parijat & Evangelista, Daniela & Guarracino, Mario. (2017). A mixed integer programming-based global optimization framework for analyzing gene expression data. *Journal of Global Optimization*. 69. 1-18. 10.1007/s10898-017-0530-0.
- ³¹⁹ Liu, Yanping & Jiao, Yang & He, Da & Fan, Qihui & Zheng, Yu & Li, Guoqiang & Wang, Gao & Yao, Jingru & Chen, Guo & Lou, Silong & Shuai, JianWei & Liu, Liyu. (2021). Deriving time-varying cellular motility parameters via wavelet analysis. *Physical biology*. 18. 10.1088/1478-3975/abfcad.
- ³²⁰ Koller D, Friedman N (2009) Probabilistic graphical models principles and techniques. MIT press, Cambridge
- ³²¹ Sugiyama, Masashi. (2016). Nonlinear Dimensionality Reduction. 10.1016/B978-0-12-802121-7.00047-9.
- ³²² Sadeghi, Jonathan. (2020). Uncertainty modelling for scarce and imprecise data in engineering applications. 10.17638/03089368.
- ³²³ Vargas, Olivia & Pérez-Ramírez, Carlos & Valtierra-Rodríguez, Martin & Yáñez-Borjas, Jesus & Amezquita-Sanchez, J.. (2021). An Explainable Machine Learning Approach Based on Statistical Indexes and SVM for Stress Detection in Automobile Drivers Using Electromyographic Signals. *Sensors*. 21. 3155. 10.3390/s21093155.
- ³²⁴ Moore, Ryan & Archer, Kristin & Choi, Leena. (2021). Statistical and machine learning models for classification of human wear and delivery days in accelerometry data. 10.1101/2020.12.31.424867.
- ³²⁵ Chapelle O, Schölkopf B, Zien A (2010) Semi-supervised learning. MIT Press, Cambridge
- ³²⁶ Khodaei, Amin & Feizi Derakhshi, Mohammad Reza & Mozaffari Tazehkand, Behzad. (2021). A Markov chain-based feature extraction method for classification and identification of cancerous DNA sequences. *BioImpacts* : BI. 11. 87-99. 10.34172/bi.2021.16.
- ³²⁷ Janssens, Hilde & Hou, Shuling & Jaeger, Johannes & Kim, Ah-Ram & Myasnikova, Ekaterina & Sharp, David & Reinitz, John. (2006). Quantitative and predictive model of

transcriptional control of the *Drosophila melanogaster* even skipped gene. *Nature genetics*. 38. 1159-65. 10.1038/ng1886.

³²⁸ Desterke, Christophe & Chiappini, Franck. (2019). Lipid Related Genes Altered in NASH Connect Inflammation in Liver Pathogenesis Progression to HCC: A Canonical Pathway. *International Journal of Molecular Sciences*. 20. 5594. 10.3390/ijms20225594.

³²⁹ Roche, Mathieu. (2011). Fouille de Textes : de l'extraction des descripteurs linguistiques à leur induction.

³³⁰ Garg, Sahil & Galstyan, Aram & Hermjakob, Ulf & Marcu, Daniel. (2015). Extracting Biomolecular Interactions Using Semantic Parsing of Biomedical Text.

³³¹ Galdón, Salvador & Pereira, Cecile & Conesa, Ana. (2020). Padhoc: a computational pipeline for pathway reconstruction on the fly. *Bioinformatics*. 36. i795-i803. 10.1093/bioinformatics/btaa811.

³³² Zerva, Chryssa & Batista-Navarro, Riza & Day, Philip & Ananiadou, Sophia. (2017). Using uncertainty to link and rank evidence from biomedical literature for model curation. *Bioinformatics (Oxford, England)*. 33. 10.1093/bioinformatics/btx466.

³³³ Usié Chimenos, Ana & Karathia, Hiren & Teixidó, Ivan & Solsona, Francesc & Alves, Rui. (2014). Biblio-MetReS for user-friendly mining of genes and biological processes in scientific documents. 10.7287/peerj.preprints.232v1.

³³⁴ Plake C, Schiemann T, Pankalla M, Hakenberg J, Leser U (2006) ALIBABA: PubMed as a graph. *Bioinformatics* 22(19):2444–2445

³³⁵ Bonner, Stephen & Barrett, Ian & Ye, Cheng & Swiers, Rowan & Engkvist, Ola & Hamilton, William. (2021). A Review of Biomedical Datasets Relating to Drug Discovery: A Knowledge Graph Perspective.

³³⁶ Soto, Axel & Zerva, Chryssa & Batista-Navarro, Riza & Ananiadou, Sophia. (2017). LitPathExplorer: A Confidence-based Visual Text Analytics Tool for Exploring Literature-Enriched Pathway Models. *Bioinformatics (Oxford, England)*. 34. 10.1093/bioinformatics/btx774.

³³⁷ Aterido, Adrià & Cañete, Juan & Tornero, Jesús & Blanco, Francisco & Fernández-Gutierrez, Benjamín & Pérez, Carolina & Alperi, Mercedes & Olivé, Alejandro & Corominas, Hèctor & Martínez-Taboada, Víctor & González-Álvaro, Isidoro & Fernández-Nebro, Antonio & Erra, Alba & López-Lasanta, María & Corbeto, Mireia & Palau, Núria & Marsal, Sara & Julià, Antonio. (2019). A Combined Transcriptomic and Genomic Analysis Identifies a Gene Signature Associated With the Response to Anti-

TNF Therapy in Rheumatoid ArthritisData_Sheet_1.pdf. *Frontiers in Immunology*. 10. 10.3389/fimmu.2019.01459.

³³⁸ DAVID: database for annotation. Visualization, and integrated discovery.

³³⁹ Canzler, Sebastian & Hackermüller, Jörg. (2020). multiGSEA: a GSEA-based pathway enrichment analysis for multi-omics data. *BMC Bioinformatics*. 21. 10.1186/s12859-020-03910-x.

³⁴⁰ Pottie, Lore & Gool, Wouter & Vanhooydonck, Michiel & Hanisch, Franz-Georg & Goeminne, Geert & Rajkovic, Andreja & Coucke, Paul & Sips, Patrick & Callewaert, Bert. (2021). Loss of zebrafish *atp6v1e1b*, encoding a subunit of vacuolar ATPase, recapitulates human ARCL type 2C syndrome and identifies multiple pathobiological signatures. *PLOS Genetics*. 17. e1009603. 10.1371/journal.pgen.1009603.

³⁴¹ Tambassi, Timothy. (2021). Ontological Categories for Geo-Ontologies. 10.1007/978-3-030-78145-3_7.

³⁴² Klapper, Isaac & Szyld, Daniel & Zhao, Kai. (2021). Metabolic Networks, Elementary Flux Modes, and Polyhedral Cones. 10.1137/1.9781611976533.

³⁴³ Guil-Asensio, Francisco & Hidalgo, Jose & Garcia, Jose. (2020). Flux Coupling and the Objective Functions' Length in EFMs. *Metabolites*. 10. 489. 10.3390/metabo10120489.

³⁴⁴ Benstead-Hume, Graeme & Chen, Xiangrong & Hopkins, Suzanna & Lane, Karen & Downs, Jessica & Pearl, Frances. (2019). Predicting synthetic lethal interactions using conserved patterns in protein interaction networks. *PLOS Computational Biology*. 15. e1006888. 10.1371/journal.pcbi.1006888.

³⁴⁵ Bademci, Guney & Edwards, Yvonne & Beecham, Gary & Khuri, Sawsan & Tekin, Demet & Martin, Eden & Scott, William & Jiang, Zhijie & Mash, Deborah & French-Mullen, Jarlath & Pericak-Vance, Margaret & Tsinoremas, Nicholas & Vance, Jeff. (2011). Pathway Analysis for Parkinson Disease: An Integrative Systems Biology Approach. A22-A22.

³⁴⁶ Ho, Hsiang-Ling & Chou, Teh-Ying & Yang, Shung-Haur & Jiang, Jeng-Kai & Chen, Wei-Shone & Chao, Yee & Teng, Hao-Wei. (2019). PD-L1 is a double-edged sword in colorectal cancer: the prognostic value of PD-L1 depends on the cell type expressing PD-L1. *Journal of Cancer Research and Clinical Oncology*. 145. 10.1007/s00432-019-02942-y.

³⁴⁷ Yi, Lim & Chin, Tan & Mohamad, Mohd & bin deris, Safaai & Subair, Saad & Ibrahim, Zuwairie. (2015). A Review on Metabolic Pathway Analysis in Biological Production. *Mini-Reviews in Organic Chemistry*. 13. 1-1. 10.2174/1570193X13666151218191358.

³⁴⁸ Pham, Vu & Liu, Lin & Bracken, Cameron & Goodall, Gregory & Li, Jiuyong & le, Thuc. (2021). Computational methods for cancer driver discovery: A survey. *Theranostics*. 11. 5553-5568. 10.7150/thno.52670.