



## LA IMPORTANCIA DE LA BIOINFORMÁTICA Y LAS TÉCNICAS ÓMICAS EN LA INDUSTRIA FARMACÉUTICA HOY EN DÍA

*Han pasado muchos años desde que a finales de la década de los 60 del pasado siglo, Margaret Dayhoff aplicara por primera vez el uso de algoritmos matemáticos usando computadoras para estudiar la historia evolutiva de las secuencias de proteínas, fundando los cimientos de la futura bioinformática o biología computacional. Seguramente ni ella misma llegó a imaginar la importancia que adquiriría en el futuro la bioinformática para el desarrollo de la medicina personalizada y su importancia en el desarrollo de la metodología de desarrollo de un fármaco a partir de una proteína determinada.*

**Antonio Gómez**, Senior Manager Data Science, PharmaLex USA

### LA LLEGADA DE LAS TECNOLOGÍAS ÓMICAS

En la década de los 90, la bioinformática empezó a estar presente en la industria farmacéutica ayudando al desarrollo de fármacos mediante estudios de interacción de proteína-ligando [1], estudiando la dinámica molecular de las interacciones de la proteína con fármacos (compuestos químicos). Es sin embargo a partir del advenimiento del siglo XXI, justo cuando los resultados de la secuenciación del genoma humano se hacen públicos [2,3], cuando la bioinformática da un paso de gigante como pilar en el análisis de la información contenida en el genoma. Gracias a este estudio sin parangón en la historia de la ciencia, se pudo tener acceso a toda la información genética recogida en el ADN (ácido desoxirribonucleico) del ser humano, que comprendía unos 3,2 Gigabytes de información que evidentemente abría un abanico de posibilidades inmenso en estudios médicos, pero a la vez implicaba formularse una pregunta fundamental: ¿Cómo podíamos conocer las implicaciones biológicas de todo ese cúmulo ingente de información

secuenciado? Es decir, se tenía el “mapa” de los genes comprendidos en el genoma humano por primera vez en la historia, pero todavía no se sabía cómo se expresaban los genes, cuáles eran sus interacciones, cómo se regulaban, etc... Es en este momento cuando asistimos al advenimiento de las llamadas tecnologías ómicas [4], que derivan de diferentes aspectos de información del genoma: transcriptómica (estudio de la expresión de los genes) [5], genómica (estudio de polimorfismos genéticos) [6], proteómica (estudio de la expresión de las proteínas) [7], metilómica (estudio de las marcas de metilación de los genes y que son importantes para su expresión) [8], etc. A partir de aquí es cuando empezamos a hablar de “Big Data” y secuenciación masiva de ADN y/o ARN.

En un primer momento, la aparición de la tecnología de microarrays [9], que ya se utilizó previamente a la secuenciación del genoma humano en la secuenciación inicial del genoma de levadura en 1995 [10], permitió por primera vez estudiar la expresión de todos los genes contenidos en el genoma humano a la vez. Esta

tecnología permite extraer ADN/ARN de una muestra biológica (fluido, biopsia) e hibridarla en una superficie sólida (chip) donde se tienen las sondas génicas correspondientes a todos los genes del genoma que están marcadas con un fluorocromo que produce una cantidad de luz al aparearse la secuencia génica de la muestra con estas sondas, la cual identifica si ese gen se expresa en esa muestra biológica y a la vez puede ser medida por una computadora que procesa esa señal en números que posteriormente se pueden analizar. Por primera vez se podían extraer muestras de un número determinado de pacientes y compararlas con controles para determinar qué genes estaban silenciados en una enfermedad en particular y cuáles estaban activados. El análisis por tanto de estos datos es crucial y la importancia de una nueva generación de profesionales en el campo de la Bioinformática, fundamental. Las tecnologías de secuenciación del genoma completo [11] poco a poco se fueron imponiendo sobre las tecnologías de microarrays; aquí, después de extraer la muestra, podemos secuenciar el ADN presente después de amplificar el material genético presente para posteriormente secuenciarlo y obtener fragmentos o lecturas desde el secuenciador que posteriormente se alinearán al genoma mediante algoritmos matemáticos para determinar la expresión de ese gen (transcriptómica) o la localización de variantes genéticas (genómica), como los dos ejemplos más claros. Estas técnicas de secuenciación masiva revolucionaron la medicina personalizada [12], así como el incremento de datos que se analizan posteriormente. Cómo se ha señalado anteriormente, la industria farmacéutica introdujo en sus pipelines de desarrollo de fármacos.

No es el objetivo de este artículo ahondar en todas las tecnologías ómicas que se utilizan para la búsqueda de nuevos fármacos, por lo que pondremos foco a los más utilizados en la industria farmacéutica actualmente.

Los estudios de asociación genómica o GWAS (en inglés) y la secuenciación del transcriptoma, permiten el descubrimiento y/o validación de nuevas dianas de

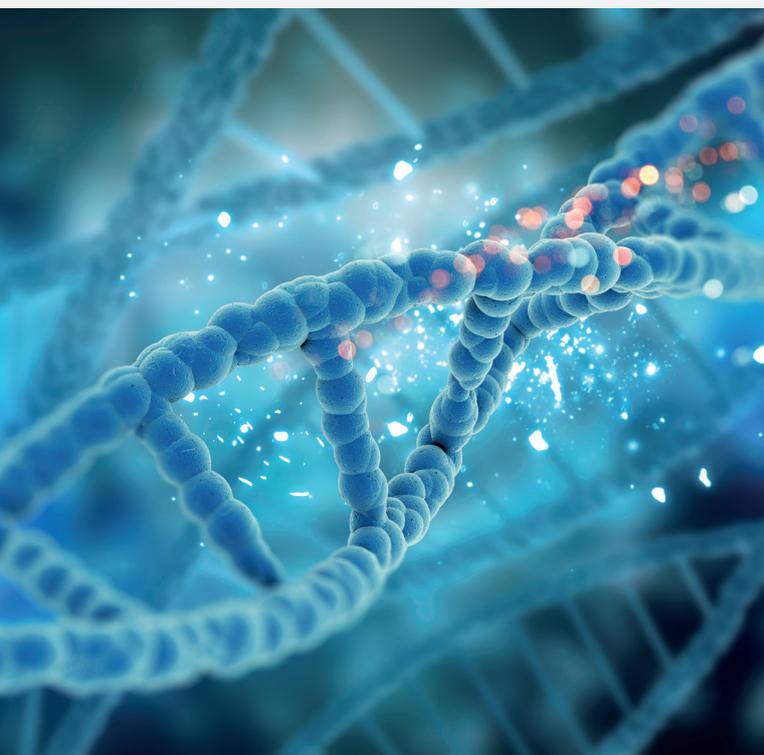
fármacos, así como validar su eficacia terapéutica y sus efectos secundarios. Una diana farmacológica, es normalmente una proteína que está intrínsecamente asociada con una enfermedad (está relacionada con una vía biológica crucial para esa enfermedad) y que puede ser accesible por un fármaco para producir una terapia deseada [13]. Las tecnologías ómicas tienen un papel importantísimo en la detección de novo de estas dianas: los estudios de asociación genómica o GWAS mediante microarrays de genotipado (hibridación de ADN de la muestra sobre microarrays con sondas de polimorfismos genéticos o SNPs) [14] o secuenciación masiva del genoma [15], permiten ayudar a su búsqueda en las primeras fases utilizando la información de variantes genéticas en una muestra muy amplia de población a estudiar. El resultado final es determinar la asociación de estas variantes con un carácter de riesgo a esa enfermedad. El objetivo es encontrar variantes genéticas que participan en el riesgo de esa enfermedad en cuestión y posteriormente localizar los genes cercanos a esa variante de riesgo. De esta manera, estos análisis sobre gran muestra poblacional pueden ayudar a identificar genes como posibles dianas de fármacos.

Estos estudios, per se, no identifican cómo las variantes afectan a otras vías posteriormente y dan lugar a la enfermedad, por lo que es necesario combinar estos resultados con información proveniente de otras ómicas, como por ejemplo la transcriptómica: identificación de las variantes asociadas a esa enfermedad que afectan a la expresión de genes implicados en la misma, estudios eQTL, expresión génica asociada a un locus de rango cuantitativo [16] o la proteómica: identificación de las variantes

asociadas a esa enfermedad que afectan a la expresión de las proteínas implicadas en la misma, estudios pQTL [17], expresión proteica asociada a un locus de rango cuantitativo, . Los estudios de asociación genómica finalmente son de importancia cabal en la farmacogenómica, cómo una variante genética de terminada afecta a un tratamiento [18], es decir, cómo un mismo tratamiento por un fármaco tiene diferentes resultados



*“Las tecnologías ómicas son la base actual del desarrollo de la farmacogenómica”*



en pacientes que padecen la misma enfermedad porque afecte a la respuesta del fármaco de manera diferente.

La transcriptómica, el análisis de la expresión de los genes en el genoma mediante técnicas de secuenciación de RNA (RNAseq) [19], permite determinar qué genes cambian en una enfermedad determinada comparando pacientes contra controles y también es útil en estudios clínicos al comparar cómo cambia la expresión de los genes al aplicar un fármaco en diferentes dosis. Estos estudios permiten determinar biomarcadores, genes que son marcadores de esa enfermedad, que ayuden a entender su funcionamiento e implicación en vías biológicas de la misma y su potencial uso como diana farmacológica. Es importante señalar que en la última década, la transcriptómica ha evolucionado hasta el punto de que es posible extraer el material genético de una sola célula, scRNAseq (secuenciación de RNA de célula individual) [20], lo que permite ir más allá en la investigación de marcadores, especialmente en oncología al comparar poblaciones de células resistentes contra células sensibles de un tumor en respuesta a un fármaco y poder ver cuáles son los cambios a nivel transcriptómico de esas células resistentes y ver en qué difieren respecto a las sensibles.

En definitiva, la revolución que las tecnologías ómicas han traído a la industria farmacéutica y al desarrollo de la medicina personalizada, ha influido en la generación de nuevos fármacos y en un cambio de paradigma

más centrado a cambios en el genoma a la hora de la búsqueda de dianas. El peso de esta tecnología y de los profesionales (Bioinformáticos, Bioestadísticos,) en la industria cada vez es mayor, debido a la cantidad de datos generados y a la necesidad de unos análisis precisos para obtener unos resultados satisfactorios. Las tecnologías ómicas son la base actual del desarrollo de la farmacogenómica, evolucionan y a la vez evolucionan los métodos matemáticos y computacionales que están permitiendo desarrollar unos fármacos más precisos y efectivos a medida que pasa el tiempo ●

#### Referencias

1. Lybrand, TP (1995) "Ligand-protein docking and rational drug design" *Curr Opin Struct Biol* ;5(2):224-8
2. Venter JC, et al. (2001). "The sequence of the human genome". *Science*. 291 (5507): 1304–51.
3. Lander, E. S. et al. (2001) "Initial sequencing and analysis of the human genome". *Nature* (409), 860–921
4. Schneider MV and Orchard S (2011) "Omics technologies, data and bioinformatics principles" *Methods Mol Biol*; 719:3-30.
5. Lowe R et al (2017) "Transcriptomics technologies" *PLoS Comput Biol*. 18 ;13(5):e1005457.
6. Uffelmann E (2021) "Genome-wide association studies" *Nat Rev Methods Primers*. 1(59)
7. Timp W (2020) "Beyond mass spectrometry, the next step in proteomics" *Sci Adv*. 6 (2)
8. Li et al (2021) "DNA methylation methods: Global DNA methylation and methylomic analyses" *Methods* (197) 28-43
9. Gresham D et al (2008) "Comparing whole genomes using DNA microarrays" *Nat Rev Genet*. (9) 291-302
10. Bussey H et al (1995) The nucleotide sequence of chromosome I from *Saccharomyces cerevisiae*". *Proc Natl Acad Sci U S A* (25) 3809-3813
11. Ng, PC and Kirkness EF (2010) "Whole genome sequencing" *Methods Mol Biol* .(628):215-26.
12. Gonzalez-Garay ML (2014) "The road from next-generation sequencing to personalized Medicine" *Per Med*. 11(5): 523–544.
13. Emmerich CH et al (2021) "Improving target assessment in biomedical research: the GOT-IT recommendations" *Nat Rev Drug Discov*. (20) 64-81
14. Joost AM et al (2021) "A comparison of genotyping arrays" *Eur J Hum Genet*. (29) 1611-1624
15. Suwinski P et al (2019) "Advancing Personalized Medicine Through the Application of Whole Exome Sequencing and Big Data Analytics" *Front. Genet* (12) 10-49.
16. Nica CA and Dermitzakis ET (2013) "Expression quantitative trait loci: present and future" *Philos Trans R Soc Lond B Biol Sci*. 368(1620): 20120362
17. Pietzner et al. (2021) "Mapping the proteo-genomic convergence of human diseases" *Science* (374) 6569
18. McInnes C. et al. (2021) "Genomewide Association Studies in Pharmacogenomics" *Clin Pharmacol Ther* 110(3):637-648
19. Wang Z. et al (2009) "RNA-Seq: a revolutionary tool for transcriptomics" *IO*(1):57-63
20. Hwang B. et al. (2018) "Single-cell RNA sequencing technologies and bioinformatics pipelines" *Exp Mol Med*. (50) 1-14