



REDEFINIENDO EL ESTUDIO DEL GENOMA: LONG-READ SEQUENCING

El descubrimiento de Fred Sanger revolucionó la biología molecular al proporcionar una forma de determinar el orden exacto de los nucleótidos en una molécula de ADN. A pesar de ser altamente laboriosa, la secuenciación de Sanger, con su alta precisión, fue crucial para crear un mapa detallado del ADN y para secuenciar casi en su totalidad el genoma humano en el Proyecto Genoma Humano.



GRETA CARMONA ANTOÑANZAS,
Directora de I+D en Health in Code.

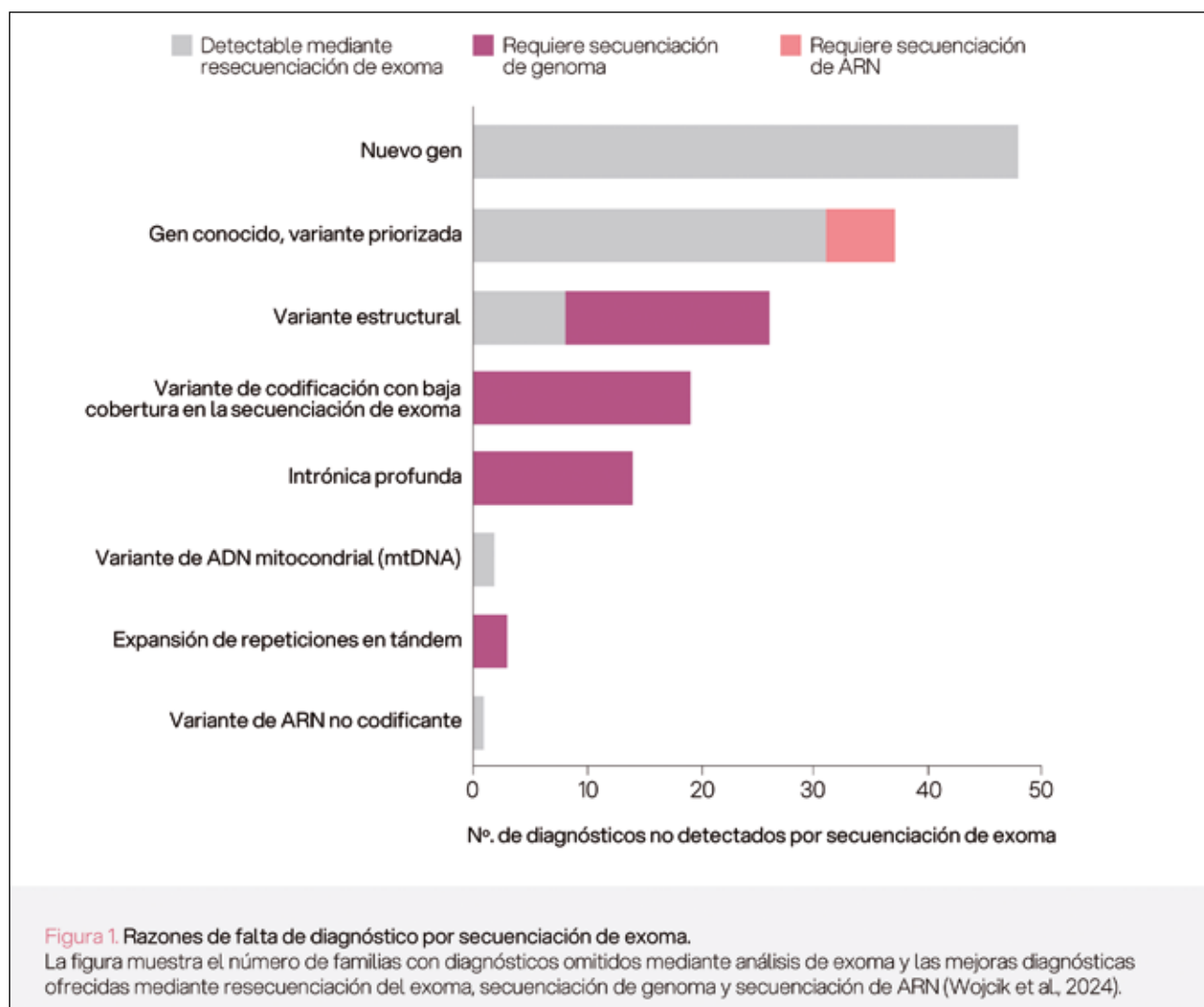
Desde su finalización, varios factores han impulsado el uso más amplio de la genética en la medicina, destacando el aumento de interés público por la salud, así como la reducción de costes y el avance de las tecnologías de secuenciación, que han hecho que la genética sea más accesible para una variedad de aplicaciones.

Entre ellas, la secuenciación masiva ha revolucionado la genómica al permitir la obtención rápida y a gran escala de datos genéticos a un coste mucho menor que en el pasado. Antes, secuenciar largas fracciones del genoma era una tarea que requería años de trabajo y enormes inversiones financieras, pero con las tecnologías de secuenciación de nueva generación (NGS), es posible analizar millones de fragmentos de ADN simultáneamente. Esto ha acelerado el descubrimiento de variantes genéticas asociadas con enfermedades, permitiendo identificar mutaciones y cambios en el ADN que contribuyen a diversas condiciones médicas. La secuenciación masiva ha llevado a la creación de bases de datos genómicas globales, que facilitan el análisis de asociación entre variaciones genéticas y enfermedades, mejorando las estrategias de salud pública al proporcionar una comprensión más profunda de las enfermedades y sus factores genéticos.

Actualmente, la genómica está integrada en prácticamente todas las especialidades médicas, proporcionando

herramientas fundamentales para el diagnóstico, tratamiento y prevención de enfermedades. En oncología, por ejemplo, la identificación de mutaciones específicas en tumores permite desarrollar terapias dirigidas más efectivas y personalizadas, mientras que, en otras especialidades, la genética desempeña un papel crucial al ayudar en el diagnóstico y manejo de condiciones hereditarias que afectan tanto a adultos como a niños (Van El et al., 2013). Podríamos decir que, la genómica está redefiniendo la práctica médica, pasando de tratamientos estandarizados a enfoques personalizados basados en la información genética de cada individuo.

El enfoque de la biología solía estar en el genoma humano en su conjunto, pero ahora la atención se centra en la variación genética individual. Este cambio hacia el estudio de la variación genética individual no solo democratiza la biología, permitiendo a las personas obtener una visión detallada de su propio perfil genético, sino que también destaca la importancia del análisis del genoma completo (WGS) en comparación con el estudio del exoma. Aunque el análisis del exoma ha supuesto un importante avance en el diagnóstico de enfermedades de origen genético, esta estrategia se centra en las regiones codificantes del ADN, mientras que el genoma completo ofrece una visión más integral al incluir tanto las regiones codificantes como las



no codificantes, que también juegan un papel crucial en la regulación genética y la salud.

Un estudio reciente, publicado en *New England Journal of Medicine* (Wojcik et al., 2024), mostró la capacidad mejorada de la secuenciación de genoma frente al exoma, concluyendo que la secuenciación mediante WGS permite el diagnóstico de un 8% de pacientes adicionales a la secuenciación de exomas (Figura 1). Este método es especialmente útil cuando las variantes genéticas responsables de una enfermedad no se encuentran en las regiones codificantes del ADN, que es lo que analiza el exoma, sino en regiones no codificantes o intrónicas profundas, las cuales solo pueden ser detectadas mediante la secuenciación del genoma completo. Los análisis de exoma y genoma completo mediante tecnologías de lecturas cortas indican una elevada capacidad diagnóstica, sin embargo, únicamente permitieron identificar variantes causales en el 30% de familias con sospecha de patologías genéticas. Esta limitación puede deberse a la presencia de variantes complejas, regiones especialmente repetitivas, variantes

estructurales, expansiones de repeticiones en tándem y variantes no codificantes intrónicas profundas que no pueden detectarse de manera fiable mediante la batería de análisis genéticos disponibles actualmente en laboratorios de diagnóstico clínico.

Impulsada por la necesidad de superar estas restricciones, han surgido varias aproximaciones de secuenciación de tercera generación que permiten analizar regiones del genoma en tiempo real produciendo lecturas completas de hasta 100 veces más largas que las tecnologías de lectura corta. La secuenciación de lecturas largas está entrando en la práctica clínica permitiendo analizar segmentos que oscilan entre 10 kb y 100 kb. Esto facilita significativamente el ensamblaje y alineamiento de las regiones de ADN con respecto al genoma de referencia, mejorando notablemente la resolución y permitiendo una correcta asociación de estas regiones con su ubicación adecuada.

Entre los mayores desafíos técnicos en su resolución, está la detección de reordenamientos cromosómicos. En patologías como el cáncer, la secuenciación de lecturas

El avance tecnológico en la secuenciación masiva ha revolucionado el campo de la genómica, proporcionando herramientas que permiten alcanzar un diagnóstico genético de manera más rápida y precisa que nunca.

largas ha sido crucial para entender cómo las diferentes variantes del oncogén de fusión BCR-ABL1 pueden afectar la respuesta al tratamiento (Cavelier et al., 2015). En pacientes con patologías raras no diagnosticadas, nos permitirá detectar reordenamientos aparentemente equilibrados como inversiones o translocaciones que no afectan al número de copias genéticas en los portadores, pero podrían tener serias repercusiones clínicas. En neuropatías con alta incidencia de expansiones de repetición, como el síndrome de CANVAS, que afecta a más de 100 nuevos casos cada año (Cortese et al., 2020), o la contracción del alelo *D4Z4*, asociado a la atrofia muscular en cara, hombros y brazos, las técnicas actuales conllevan procesos diagnósticos altamente costosos y largos en el tiempo, desencadenando la frustración de pacientes y sus familias tediosas. Más del 10% del genoma está constituido por pseudogenes (Lughmani et al., 2024), resultado de duplicaciones masivas de genes durante procesos evolutivos que luego pierden su función y no se diferencian analíticamente de sus homólogos activos mediante lecturas cortas. Por ejemplo, la alta homología en genes como *SMN1* y *SMN2* impide el diagnóstico preciso de la atrofia muscular espinal, mientras que en el caso de la distrofia macular de Stargardt (Whelan et al., 2023), causa más frecuente de pérdida de visión central, se diagnostica de manera tardía debido a la presencia de múltiples variantes intrónicas profundas que no son fácilmente detectables con otras técnicas de análisis molecular.

El avance tecnológico en la secuenciación masiva ha revolucionado el campo de la genómica, proporcionando herramientas que permiten alcanzar un diagnóstico genético de manera más rápida y precisa que nunca. Al realizar estudios genómicos, se generan grandes volúmenes de datos sobre variantes genéticas y sus asociaciones con condiciones específicas. Una base de datos rica y

actualizada no solo permitirá a los profesionales de la salud acceder a información relevante sobre variantes genéticas y sus implicaciones clínicas, sino que también promoverá el aprendizaje colaborativo entre investigadores y clínicos (Rhie et al., 2021). Esto, a su vez, impulsará el aumento del conocimiento sobre enfermedades raras y genéticas, contribuyendo a diagnósticos más certeros y oportunos. En conjunto, la combinación de avances en secuenciación, la selección adecuada de técnicas diagnósticas y la mejora de bases de datos clínicas son elementos clave para avanzar hacia un futuro en el que el diagnóstico y tratamiento de enfermedades genéticas sean más precisos, y personalizados.

Bibliografía

- Cavelier, L., Ameer, A., Häggqvist, S., Höjjer, I., Cahill, N., Olsson-Strömberg, U., & Hermanson, M. (2015). Clonal distribution of BCR-ABL1 mutations and splice isoforms by single-molecule long-read RNA sequencing. *BMC Cancer*, 15(1). <https://doi.org/10.1186/S12885-015-1046-Y>.
- Cortese, A., Reilly, M. M., & Houlden, H. (2020). RFC1 CANVAS / Spectrum Disorder. *GeneReviews*. <https://www.ncbi.nlm.nih.gov/books/NBK564656/>.
- Lughmani, H., Patel, H., & Chakravarti, R. (2024). Structural Features and Physiological Associations of Human 14-3-3 ζ Pseudogenes. *Genes*, 15(4). <https://doi.org/10.3390/GENES15040399>.
- Rhie, A., McCarthy, S. A., Fedrigo, O., Damas, J., Formenti, G., Koren, S., Uliano-Silva, M., Chow, W., Fungtammasan, A., Kim, J., Lee, C., Ko, B. J., Chaisson, M., Gedman, G. L., Cantin, L. J., Thibaud-Nissen, F., Haggerty, L., Bista, I., Smith, M., ... Jarvis, E. D. (2021). Towards complete and error-free genome assemblies of all vertebrate species. *Nature* 2021 592:7856, 592(7856), 737–746. <https://doi.org/10.1038/s41586-021-03451-0>.
- Van El, C. G., Cornel, M. C., Borry, P., Hastings, R. J., Fellmann, F., Hodgson, S. V., Howard, H. C., Cambon-Thomsen, A., Knoppers, B. M., Meijers-Heijboer, H., Scheffer, H., Tranebjaerg, L., Dondorp, W., & De Wert, G. M. W. R. (2013). Whole-genome sequencing in health care. *European Journal of Human Genetics* 2013 21:6, 21(6), 580–584. <https://doi.org/10.1038/ejhg.2013.46>.
- Whelan, L., Dockery, A., Stephenson, K. A. J., Zhu, J., Kopčić, E., Post, I. J. M., Khan, M., Corradi, Z., Wynne, N., O'Byrne, J. J., Duignan, E., Silvestri, G., Roosing, S., Cremers, F. P. M., Keegan, D. J., Kenna, P. F., & Farrar, G. J. (2023). Detailed analysis of an enriched deep intronic ABCA4 variant in Irish Stargardt disease patients. *Scientific Reports*, 13(1). <https://doi.org/10.1038/S41598-023-35889-9>.
- Wojcik, M. H., Lemire, G., Berger, E., Zaki, M. S., Wissmann, M., Win, W., White, S. M., Weisburd, B., Wieczorek, D., Waddell, L. B., Verboon, J. M., VanNoy, G. E., Töpf, A., Tan, T. Y., Syrbe, S., Strehlow, V., Straub, V., Stenton, S. L., Snow, H., ... O'Donnell-Luria, A. (2024). Genome Sequencing for Diagnosing Rare Diseases. *New England Journal of Medicine*, 390(21), 1985–1997. https://doi.org/10.1056/NEJM0A2314761/SUPPL_FILE/NEJM0A2314761_DATA-SHARING.PDF.