

TECNOLOGÍAS PARA LA DETECCIÓN DE VARIANTES FARMACOGENÉTICAS

El campo de la farmacogenómica (PGx) se está desarrollando rápidamente. En los últimos 20 años, no solo ha habido un gran progreso en el desarrollo de las pautas de PGx, también ha habido un avance tecnológico significativo que ha aportado nuevas tecnologías para evaluar variantes genéticas. En PGx, este tipo de variación son fundamentalmente variaciones puntuales o SNV (Single Nucleotide Variants) y en menor medida pequeñas inserciones y deleciones, aunque algunos genes también presentan variaciones en el número de copias de interés clínico. En este documento, revisaremos las tecnologías más utilizadas para analizar todos los tipos de variación genética que podemos encontrarnos en el diagnóstico PGx

1. Secuenciación Sanger

La secuenciación Sanger consiste en determinar la composición de los nucleótidos que conforman un fragmento diana de ADN amplificado mediante la reacción en cadena de la polimerasa (PCR). Se basa en la síntesis de forma secuencial de la cadena complementaria de una de las hebras del fragmento diana (ADN molde) usando el método enzimático Sanger en el que participa la Taq polimerasa, un cebador y los didesoxinucleótidos marcados con fluorescencia (ddNTPs: ddATP, ddCTP, ddGTP, ddTTP), cada uno marcado con un fluoróforo distinto. De este modo, se produce una "escalera" de ADN de fragmentos que difieren en una base de longitud. Posteriormente, cada fragmento marcado se separará por tamaño por electroforesis capilar, con detección automatizada de los fragmentos de ADN marcados fluorescentemente, proporcionando la secuencia ordenada de los fragmentos en cromatogramas. La secuencia del ADN inicial a partir de los fragmentos de ADN secuenciados se deduce bioinformáticamente por análisis de secuencias solapadas de los fragmentos de ADN ¹

2. MLPA (multiplex ligation-dependent probe amplification)

MLPA (amplificación de sondas tras ligación múltiple) es un método basado en PCR multiplex utilizado para detectar alteraciones en el número de copias en regiones concretas y permite la amplificación de hasta 40 regiones de ADN. Utiliza sondas de

dos oligonucleótidos que reconocen sitios adyacentes en el ADN, y que contienen una secuencia universal que permite la amplificación simultánea por PCR de todas las sondas con sólo una pareja de cebadores². Además, uno de los oligonucleótidos contiene una secuencia de longitud variable (secuencia stuffer), que permite la separación de los fragmentos amplificados durante la electroforesis. El MLPA presenta relativa simplicidad, bajo coste y posibilidad de realizar la técnica en cualquier laboratorio ya que sólo se necesita un termociclador y un secuenciador capilar.²

3. Ensayos de genotipificado *singleplex*

Estos ensayos permiten llevar a cabo la determinación de sustituciones puntuales (SNVs) y pequeñas inserciones/deleciones (InDels). Estos ensayos nos permiten llevar a cabo una discriminación alélica mediante una amplificación con PCR a tiempo real. En ella se utilizan una pareja de cebadores de PCR preoptimizados y dos sondas alelo-específicas, una con marcaje del fluorocromo FAM y otra con marcaje del fluorocromo VIC en el extremo 5' y uniones al surco menor (MGB) y supresores no fluorescentes (*Nonfluorescent quencher (NFQ)* en el extremo 3' (TaqMan®). La actividad de exonucleasa de la ADN polimerasa AmpliTaq Gold® fragmenta únicamente las sondas hibridadas y esta fragmentación separa el fluorocromo indicador (VIC o FAM) del supresor (quencher), lo que permite la emisión de la fluorescencia del indicador y por lo tanto, la señal de fluorescencia generada por la amplificación de la PCR indica los alelos que están presentes en la muestra³. Otro tipo de ensayos muy utilizados son los ensayos basados en la amplificación competitiva de alelos específicos por PCR de las secuencias objetivo y el genotipificado de fluorescencia de punto final (KASPar™) utilizando un lector de fluorescencia final⁴. Este tipo de aproximación también pueden ser aplicada a la determinación del número de copias de nuestra región de interés, ya que permite determinar el número de copias relativo de la secuencia objeto de interés en una muestra de ADN_g, normalizada según el número de copias conocido de una secuencia de referencia.⁵

4. Ensayos de genotipificado *multiplex*

La prueba de ensayos multiplex de SNVs es la tecnología más utilizada en PGx, ya sea a través de paneles comerciales o paneles personalizados. Estos paneles pueden contener un conjunto preseleccionado de SNVs, que, según la plataforma, pueden variar desde unas pocas variantes en un solo gen hasta miles de variantes en todo el genoma.

Los paneles disponibles comercialmente contienen típicamente variantes que están vinculadas a la respuesta al fármaco en las pautas de PGx o en PharmGKB^{6,7}. La evidencia subyacente a las variantes seleccionadas puede variar, desde arrays que solo contienen las variantes más fuertemente asociadas, hasta paneles que contienen todas las variantes teóricamente asociadas con la respuesta al fármaco, por ejemplo, todas aquellas dentro de los genes conocidos relacionados con el fármaco.

Estos paneles están basados en discriminación alélica mediante PCR e hibridación con sondas alelo-específicas, combinada con una forma de detección de fluorescencia o quimioluminiscencia para determinar el genotipo que está presente en el sitio de interés⁸. Otra tecnología se basa en el uso de espectrometría de masas, que permite llevar a cabo el genotipificado mediante la detección de diferencias de masa entre el nucleótido *wt* y mutantes⁹. También existen tecnologías basadas en la hibridación del ADN fragmentado en un soporte sólido o array donde se encuentran inmovilizadas una colección de sondas complementarias a cada las variantes que se pretenden interrogar^{10,11}. Esto paneles pueden incluir incluso el análisis de marcadores en genes más complejos como HLA o permitir la detección de CNVs aunque la deleciones o inserciones deben ser de gran tamaño (varias kbs) para poder ser detectadas. Por último, también existen arrays comerciales con cobertura de todo el genoma con contenido específico farmacogenético.

Esto resulta muy interesante ya que permite no solo identificar las variantes clínicamente relevantes sino generar información adicional que pueda ser útil con fines de investigación. Algunas de estas tecnologías también permiten desarrollar tu propio panel clínico personalizado, permitiendo incorporar variantes con evidencia clínica

menor que requieran de una validación clínica prospectiva, variantes específicas de población o variantes raras identificadas mediante estudios de resecuenciación.

5. Secuenciación masiva (*Next Generation Sequencing*)

El desarrollo en los últimos años de las denominadas tecnologías de secuenciación masiva permite actualmente obtener millones de secuencias de ADN a una velocidad sin precedentes y a un coste cada vez más reducido ¹². Además, la secuenciación masiva tiene el potencial de detectar todos los tipos de variación genómica, incluyendo variantes de nucleótido único o SNVs, pequeñas inserciones y deleciones, y también variantes estructurales tanto equilibradas (inversiones y traslocaciones) como desequilibradas (deleciones o duplicaciones). A pesar de existir diferentes tecnologías de secuenciación que difieren en varios aspectos, el esquema de trabajo es similar para todas ellas ¹³. En términos generales se da una fragmentación del ADN y mediante ligación se añaden secuencias adaptadoras a los extremos. Los fragmentos de ADN se amplifican clonalmente y se agrupan juntos (*clustering*) para ser utilizados como entidades a secuenciar. La secuenciación se realiza entonces alternando ciclos de terminación reversible cíclica y de toma de imágenes. Las secuencias cortas, entre ~100-500 pares de bases dependiendo de la tecnología utilizada, producidas por el instrumento a partir de los extremos del ADN con los adaptadores, se denominan lecturas o *reads*. Estas lecturas se alinean con el genoma de referencia y las variantes se identifican en función de las variaciones con la referencia. Un aspecto importante en la NGS es el número de veces que cada base del genoma está presente en los *reads* de secuenciación producidos. Este valor se denomina profundidad de cobertura (*depth of coverage*, o simplemente, *coverage*) y es uno de los factores determinantes para evaluar la fiabilidad del nucleótido asignado a esa posición del genoma. La secuenciación masiva se puede aplicar con tres enfoques diferentes: secuenciación dirigida de una región o panel de genes de interés, secuenciación de las regiones codificantes del genoma humano o exoma, que cubre aproximadamente el 1-2% de todo el genoma (*Whole Exome Sequencing*) y secuenciación del genoma completo (*Whole Genome Sequencing*). La incorporación de estas tecnologías en el campo del diagnóstico farmacogenético se ha visto aumentada en los últimos años ¹⁴⁻¹⁶ ya que esta aproximación permite tanto la identificación de variantes farmacogenéticas

clínicamente relevantes conocidas como la identificación de nuevas variantes, ayudando a contribuir sustancialmente a mejorar el proceso de diagnóstico farmacogenético y esclarecer el papel de variantes menos frecuentes en la población en la aparición de los efectos adversos. Sin embargo, a pesar de que este tipo de aproximaciones cada vez son más competitivas económicamente gracias al descenso de los precios en los últimos años, todavía existe dificultad en el procesamiento dada la gran cantidad de datos generada.

En estos últimos años se están incorporado al mercado tecnologías de secuenciación que permiten la lectura de secuencias más largas, de hasta 45 kB¹⁷. Se ha demostrado que la secuenciación de lectura larga es capaz de resolver loci complejos en todo el genoma¹⁸. En el campo de la farmacogenética, el gen más complejo es CYP2D6, que contiene tanto SNVs como variaciones estructurales. Se ha demostrado que, con la secuenciación de lectura larga, el gen CYP2D6 (~ 6,6 kb) puede secuenciarse en una lectura completa y resolverse completamente en haplotipos en fase, incluidas las variantes estructurales¹⁹. Aunque la secuenciación de lectura larga en PGx está actualmente limitada a estudios de un solo gen y no se han realizado estudios a gran escala que apliquen secuenciación de lectura larga en PGx clínica, resulta un campo muy prometedor para el análisis de genes de estructura compleja que hasta la fecha no están siendo adecuadamente caracterizados.²⁰.

6. Genotipificación del HLA

El sistema del antígeno leucocitario humano (HLA) (el complejo mayor de histocompatibilidad [CMH] en los seres humanos) es un componente importante del sistema inmunitario y está controlado por genes localizados en el cromosoma 6. La determinación del genotipo de los genes HLA es un procedimiento complejo debido al extremo grado de polimorfismo en estos genes. La genotipificación del HLA se realiza mediante métodos basados en el ADN, como SSP (*sequence-specific primer*), el SSO (*sequence-specific oligonucleotide*), y la tipificación basada en la secuencia (*sequence-based typing*; SBT)²¹. Sin embargo, los recientes avances en las tecnologías de secuenciación masiva han tenido un impacto significativo en el proceso de genotipificación de HLA y se han establecido diferentes métodos basados en NGS, como secuenciación basada en amplicones²² o y la genotipificación a través de

secuenciación del exoma o genoma completo²³. También la secuenciación con lectura larga se ha empezado a aplicar a la caracterización de estos genes²⁴.

7. Secuenciación versus genotipificación

Mientras que la secuenciación nos permite identificar todos los posibles alelos que existen en una posición concreta en nuestro genoma, cabe señalar que los ensayos de genotipificación habitualmente nos permiten determinar sólo dos de las alternativas genéticas, de acuerdo al diseño de las sondas empleadas (por ejemplo, en los ensayos de discriminación alélica tipo sondas Taqman) por lo que hay que tener en cuenta esta limitación cuando se está caracterizando variantes trialélicas como por ejemplo, la variante rs28371706 en el gen CYP2D6.

Además, cabe destacar que la secuenciación (Sanger y NGS) permite el descubrimiento de variantes farmacogenéticas menos frecuentes que pueden contribuir sustancialmente a explicar la totalidad de la variabilidad interindividual fenotípica de respuesta a un fármaco; algunas pueden ser variantes deletéreas, como variantes de pérdida de función (por ejemplo, *nonsense*, mutación sin sentido) pero también podemos encontrarnos con variantes de significado clínico desconocido (*variants of unknown significance*, VUS) como, por ejemplo, aquellas variantes que conllevan un cambio de aminoácido del que se desconoce su efecto en la función de la proteína.

REFERENCES

1. Mitchelson K, Jing Cheng. *Methods in Molecular Biology. Capillary Electrophoresis of Nucleic Acids.*; 2001.
2. Eijk-Van Os PGC, Schouten JP. *Multiplex Ligation-Dependent Probe Amplification (MLPA®) for the Detection of Copy Number Variation in Genomic Sequences.* Vol 688.; 2011. doi:10.1007/978-1-60761-947-5_8
3. De Kok JB, Wiegerinck ETG, Giesendorf BAJ, Swinkels DW. Rapid genotyping of single nucleotide polymorphisms using novel minor groove binding DNA

- oligonucleotides (MGB probes). *Hum Mutat.* 2002;19(5):554-559.
doi:10.1002/humu.10076
4. Lai K, Lorenc MT, Edwards D. *Molecular Marker Databases*. Vol 1245.; 2015.
doi:10.1007/978-1-4939-1966-6_4
 5. Schaeffeler E, Schwab M, Eichelbaum M, Zanger UM. CYP2D6 Genotyping Strategy Based on Gene Copy Number Determination by TaqMan Real-Time PCR. *Hum Mutat.* 2003;22(6):476-485. doi:10.1002/humu.10280
 6. Arbitrio M, Scionti F, Di Martino MT, et al. Pharmacogenomics Biomarker Discovery and Validation for Translation in Clinical Practice. *Clin Transl Sci.* 2021;14(1):113-119. doi:10.1111/cts.12869
 7. Arbitrio M, Di Martino MT, Scionti F, et al. DMET™ (Drug Metabolism Enzymes and Transporters): A Pharmacogenomic platform for precision medicine. *Oncotarget.* 2016;7(33):54028-54050. doi:10.18632/oncotarget.9927
 8. Raso A, Biassoni R. *A Quarter Century of PCR-Applied Techniques and Their Still-Increasing Fields of Use*. Vol 2065.; 2020. doi:10.1007/978-1-4939-9833-3_1
 9. Gabriel S, Ziaugra L, Tabbaa D. SNP genotyping using the sequenom massARRAY iPLEX Platform. *Curr Protoc Hum Genet.* 2009;(SUPPL. 60):1-18. doi:10.1002/0471142905.hg0212s60
 10. Fan JB, Gunderson KL, Bibikova M, et al. [3] Illumina Universal Bead Arrays. *Methods Enzymol.* 2006;410(06):57-73. doi:10.1016/S0076-6879(06)10003-8
 11. Lamy P, Andersen CL, Wikman FP, Wiuf C. Genotyping and annotation of Affymetrix SNP arrays. *Nucleic Acids Res.* 2006;34(14):1-9. doi:10.1093/nar/gkl475
 12. McCombie WR, McPherson JD, Mardis ER. Next-generation sequencing technologies. *Cold Spring Harb Perspect Med.* 2019;9(11). doi:10.1101/cshperspect.a036798
 13. Levy SE, Myers RM. Advancements in Next-Generation Sequencing. *Annu Rev Genomics Hum Genet.* 2016;17:95-115. doi:10.1146/annurev-genom-083115-022413
 14. van der Lee M, Allard WG, Bollen S, et al. Repurposing of Diagnostic Whole Exome Sequencing Data of 1,583 Individuals for Clinical Pharmacogenetics. *Clin*

- Pharmacol Ther.* 2020;107(3):617-627. doi:10.1002/cpt.1665
15. Reisberg S, Krebs K, Lepamets M, et al. Translating genotype data of 44,000 biobank participants into clinical pharmacogenetic recommendations: challenges and solutions. *Genet Med.* 2019;21(6):1345-1354. doi:10.1038/s41436-018-0337-5
 16. Cousin MA, Matey ET, Blackburn PR, et al. Pharmacogenomic findings from clinical whole exome sequencing of diagnostic odyssey patients. *Mol Genet Genomic Med.* 2017;5(3):269-279. doi:10.1002/mgg3.283
 17. Mantere T, Kersten S, Hoischen A. Long-read sequencing emerging in medical genetics. *Front Genet.* 2019;10(MAY):1-14. doi:10.3389/fgene.2019.00426
 18. Wenger AM, Peluso P, Rowell WJ, et al. Accurate circular consensus long-read sequencing improves variant detection and assembly of a human genome. *Nat Biotechnol.* 2019;37(10):1155-1162. doi:10.1038/s41587-019-0217-9
 19. Bowden R, Davies RW, Heger A, et al. Sequencing of human genomes with nanopore technology. *Nat Commun.* 2019;10(1):1-9. doi:10.1038/s41467-019-09637-5
 20. Ingelman-Sundberg M. Translation of pharmacogenomic drug labels into the clinic. Current problems. *Pharmacol Res.* 2020;153(December 2019):104620. doi:10.1016/j.phrs.2019.104620
 21. Bontadini A. HLA techniques: Typing and antibody detection in the laboratory of immunogenetics. *Methods.* 2012;56(4):471-476. doi:10.1016/j.ymeth.2012.03.025
 22. Schöfl G, Lang K, Quenzel P, et al. 2.7 million samples genotyped for HLA by next generation sequencing: Lessons learned. *BMC Genomics.* 2017;18(1):1-16. doi:10.1186/s12864-017-3575-z
 23. Major E, Rigó K, Hague T, Bérces A, Juhos S. HLA typing from 1000 Genomes whole genome and whole exome illumina data. *PLoS One.* 2013;8(11). doi:10.1371/journal.pone.0078410
 24. Matern BM, Olieslagers TI, Groeneweg M, et al. Long-Read Nanopore Sequencing Validated for Human Leukocyte Antigen Class I Typing in Routine Diagnostics. *J Mol Diagnostics.* 2020;22(7):912-919. doi:10.1016/j.jmoldx.2020.04.001