



COMUNICACIÓN BREVE

Evaluation of linkage disequilibrium between CYP2D6 gene polymorphisms associated with breast cancer in women

Evaluación del desequilibrio de ligamiento entre los polimorfismos del gen CYP2D6 asociados a cáncer de mama en mujeres

Sergio Vladimir Flores¹  , Angel Roco-Videla²  , Raúl Aguilera-Eguía³  , Román Montaña⁴  

¹Universidad Arturo Prat. Iquique, Chile.

²Universidad Bernardo O'Higgins, Programa de Magister en Ciencias Químico-Biológicas. Santiago, Chile.

³Universidad Católica de la Santísima Concepción, Departamento de Salud Pública. Concepción, Chile.

⁴Universidad Autónoma de Chile, Facultad de Ciencias de la Salud. Santiago, Chile.

Citar como: Flores SV, Roco-Videla A, Aguilera-Eguía R, Montaña R. Evaluation of linkage disequilibrium between CYP2D6 gene polymorphisms associated with breast cancer in women. Salud, Ciencia y Tecnología. 2024; 4:904. <https://doi.org/10.56294/saludcyt2024904>

Recibido: 13-11-2023

Revisado: 26-02-2024

Aceptado: 29-04-2024

Publicado: 30-04-2024

Editor: Dr. William Castillo-González 

ABSTRACT

Introduction: the CYP2D6 gene is highly polymorphic and is involved in the metabolism of a wide variety of drugs and xenobiotics. Three allelic variants of the gene (rs3892097, rs1065852 and rs28371725) have been studied in relation to breast cancer, but the degree of genetic linkage between them is unknown.

Objective: to test whether the three breast cancer-associated SNPs are inherited en bloc.

Methods: genotypes of 280 SNPs of the CYP2D6 gene were obtained from the 1000Genomes database and linkage disequilibrium between the three breast cancer-associated SNPs and the other 277 SNPs in the gene was analysed. A threshold of $r^2 = 0,7$ was used to identify loci in linkage disequilibrium.

Results: a strong correlation was found between rs3892097 and rs1065852 in Europe and Latin America. In addition, eleven polymorphisms with a high level of linkage were identified, four of which have a non-synonymous mutation effect.

Conclusions: both polymorphisms could have a causal effect on breast cancer or that one of them is associated by linkage. Continuing to study both polymorphisms and considering the eleven new polymorphisms for analysis would be of relevance to deepen knowledge in relation to breast cancer in women.

Keywords: Population Genetics; SNP; Linkage Disequilibrium; CYP2D6 Gene; Breast Cancer.

RESUMEN

Introducción: el gen CYP2D6 es altamente polimórfico y está involucrado en el metabolismo de una gran variedad de medicamentos y xenobióticos. Se han estudiado tres variantes alélicas del gen (rs3892097, rs1065852 y rs28371725) en relación con el cáncer de mama, pero se desconoce el grado de ligamiento genético entre ellas.

Objetivo: probar si los tres SNPs asociados a cáncer de mama se heredan en bloque.

Método: se obtuvieron los genotipos de 280 SNPs del gen CYP2D6 desde la base de datos 1000Genomes y se analizó el desequilibrio de ligamiento entre los tres SNPs asociados al cáncer de mama y los otros 277 SNPs del gen. Se usó un umbral de $r^2 = 0,7$ para identificar loci en desequilibrio de ligamiento.

Resultados: se encontró una fuerte correlación entre rs3892097 y rs1065852 en Europa y Latinoamérica. Además, se identificaron once polimorfismos con un alto nivel de ligamiento, cuatro de los que tienen un efecto de mutación no sinónima.

Conclusiones: ambos polimorfismos podrían tener un efecto causal sobre el cáncer de mama o que uno de ellos está asociado por ligamiento. Continuar estudiando ambos polimorfismos y considerar para su análisis los once nuevos, tendría relevancia profundizar el conocimiento en relación al cáncer de mama en mujeres.

Palabras clave: Genética de Poblaciones; SNP; Desequilibrio de Ligamiento; Gen CYP2D6; Cáncer de Mama.

INTRODUCCIÓN

El gen CYP2D6 es sumamente polimórfico diferentes poblaciones: se conocen 141 variantes alélicas que intervienen en el metabolismo de una amplia variedad de medicamentos y xenobióticos.⁽¹⁾ Entre variantes más estudiadas se encuentran el SNP rs3892097 (CYP2D6*4), rs1065852 (CYP2D6*10) y rs28371725 (CYP2D6*41).^(2,3,4,5)

En un estudio en mujeres estadounidenses y otro en austriacas se encontró que los genotipos rs3892097 CT + TT se asocian a una recurrencia de cáncer mama más breve y una peor supervivencia libre de enfermedad en comparación con el genotipo CC.^(6,7) Por otra parte, en un estudio similar en mujer suecas se obtuvieron resultados en el sentido inverso,⁽⁸⁾ mientras que en una investigación sobre rs1065852 en mujeres tailandesas se observó una menor supervivencia y un mayor riesgo de recurrencia de cáncer en las portadoras del genotipo TT.⁽⁹⁾ Finalmente, en un estudio en Singapur se encontró que las mujeres con el genotipo TT tenían menores concentraciones de fármaco activo que las del genotipo CC durante su periodo de tratamiento.⁽¹⁰⁾ Por su parte, el SNP rs28371725 ha sido estudiado en forma conjunta con los SNPs rs3892097 y rs1065852 en población de mujeres alemanas con cáncer, quienes no presentaban el alelo ancestral en forma de homocigoto en todos los polimorfismos tenían más riesgos de recurrencia de cáncer.⁽¹¹⁾

Si bien los hallazgos descritos confluyen en la asociación estadística entre los SNPs rs3892097, rs1065852 y rs28371725 y el cáncer de mama, se desconoce el grado de ligamiento genético que existe entre ellos: un alto grado de ligamiento significaría una asociación colineal entre estos marcadores y el fenotipo, es decir, redundancia en la asociación estadística.

El análisis ligamiento genético hace referencia a la herencia; cuando dos o más loci, durante la meiosis van a parar la misma célula germinal, eludiendo la recombinación, estos se heredan simultáneamente si la célula fecunda o es fecundada. Por ello si dos o más loci, se heredan de esta manera, y coincidentemente se encuentra asociación para un mismo fenotipo, dicha asociación puede deberse al ligamiento.⁽¹²⁾

Si bien, como ha quedado dicho, se han realizado estudios en diversas poblaciones específicas, bases de datos genéticos que abarquen muestras de distintos continentes simultáneamente, como la usada en esta investigación, permiten un enfoque más amplio con relación a la variabilidad humana.⁽¹³⁾

El objeto de este estudio consiste en evaluar el desequilibrio de ligamiento entre los marcadores antes mencionados, así como el que posiblemente existe con otras variantes dentro del gen CYP2D6, todo esto con un enfoque poblacional mediante el análisis de muestras genéticas de diferentes continentes.

MÉTODOS

Diseño de investigación

Estudio cuantitativo no experimental a partir de datos secundarios (in silico)

Obtención de muestra y sus características

Se obtuvo los genotipos en fase de los 280 SNPs bi alélicos del gen CYP2D6 desde la base de datos 1000Genomes en su fase 3, la que contiene 2504 individuos pertenecientes a 26 poblaciones agrupadas en cinco macro poblaciones: África (n= 661), Asia del este (n= 504), Asia del sur (n= 489), Europa (n= 503) y Latinoamérica (n= 347). Esta base de datos es actualmente utilizada para análisis genético abarcando poblaciones de diferentes lugares del mundo de manera simultánea.^(14,15)

Metodología de análisis de datos

Se analizó el desequilibrio de ligamiento (LD; linkage disequilibrium) entre los tres SNPs CYP2D6 asociados a cáncer de mama (rs3892097, rs1065852 y rs28371725), así como entre estos y los restantes 277 SNPs del gen. Para esto, se usó el coeficiente de correlación de Spearman (r^2) (0,05;95) y las herramientas de manejo de datos VCF (variant calling format) implementada en VcfTools.⁽¹⁶⁾ Se usó un límite de $r^2= 0,7$ para identificar loci en desequilibrio de ligamiento.⁽¹⁷⁾

RESULTADOS

En la figura 1a se puede observar que no existe un alto nivel de ligamiento para el polimorfismo rs28371725 y los otros 277 SNPs del gen CYP2D6 dado que todos los valores de r^2 fueron inferiores a 0,3. Por el contrario, en las figuras 1b y 1c se puede observar que para los polimorfismos rs3892097 y rs1065852 si existe un alto nivel de ligamiento con valores de r^2 por sobre 0,7 dentro de su distribución.

En la tabla 1 se observa una fuerte correlación entre los polimorfismos rs3892097 y rs1065852 en los grupos de Europa y América donde en la mayoría de los casos el valor de ligamiento fue significativo. Se identificaron, además, once polimorfismos con un alto nivel de ligamiento, sumados a los tres polimorfismos de asociación previamente descritos, aumentando a 14 el número de marcadores candidatos a estar asociados causalmente con el cáncer de mama. Dos de estos polimorfismos (rs28371704, rs28371703) tienen un efecto de mutación no sinónima, lo que sugiere que pueden tener un impacto funcional en el gen CYP2D6.⁽¹⁸⁾



Figura 1. Desequilibrio de ligamiento entre el polimorfismo rs28371725 rs3892097 y rs1065852 del gen CYP2D6 respecto de los otros 277 SNPs en el gen

Tabla 1. Comparaciones con desequilibrio de ligamiento significativo ($r^2 > 0,7$) entre los dos SNPs del gen CYP2D6 asociados a cáncer de mama versus los otros 277 SNPs en el gen

ID1	ID2	POS1	POS2	DIF (bp)	África	Asia Este	Asia Sur	Europa	América
rs3892097	rs2004511	42524947	42523211	1736	0,492	0,001	0,619	0,881*	0,836*
rs3892097	rs58440431	42524947	42524696	251	0,458	0,001	0,614	0,909*	0,855*
rs3892097	rs28371705	42524947	42525798	851	0,324	0,249	0,705*	0,942*	0,705*
rs3892097	rs28371704	42524947	42525811	864	0,324	0,500	0,715*	0,916*	0,705*
rs3892097	rs28371703	42524947	42525821	874	0,324	0,500	0,715*	0,916*	0,705*
rs3892097	rs1065852	42524947	42526694	1747	0,507	0,001	0,623	0,903*	0,855*
rs1065852	rs2004511	42526694	42523211	3483	0,956*	0,933*	0,993*	0,976*	0,955*
rs1065852	rs58440431	42526694	42524696	1998	0,889*	0,913*	0,970*	0,982*	0,955*
rs1065852	rs1081003	42526694	42525756	938	0,786*	0,988*	0,452	0,097	0,305

rs1065852	rs28371705	42526694	42525798	896	0,164	0,000	0,439	0,850*	0,602
rs1065852	rs28371704	42526694	42525811	883	0,164	0,001	0,446	0,827*	0,603
rs1065852	rs28371703	42526694	42525821	873	0,164	0,001	0,446	0,827*	0,603

*ligamientos con $r^2 > 0,7$.

DISCUSIÓN

Los altos niveles de ligamiento encontrados para los polimorfismos rs3892097 y rs1065852, sugiere que el conjunto de polimorfismos en ligamiento estrecho en el gen CYP2D6 son marcadores candidatos a una asociación causal con el cáncer de mama. El que el polimorfismo rs28371725 no presente un alto nivel de ligamiento sugiere una asociación independiente con el cáncer de mama, lo que amerita ser estudiado. Por otra parte, en este estudio solamente se analizó el desequilibrio de ligamiento entre locis de un mismo gen y por lo tanto dentro de un mismo cromosoma, siendo importante, para una futura investigación realizar análisis de desequilibrio de ligamiento entre locis en cromosomas diferentes. Esto puede parecer contradictorio, dado que como se explico es generalmente la proximidad física aquella que determina si dos loci se heredan en bloque. Sin embargo, esto no toma en cuenta que es posible que una combinación de variantes genéticas en diferentes cromosomas puede resultar letal para el cigoto, embrión o feto en desarrollo, lo que se convierte en factor de selección natural y amplía el concepto de desequilibrio de ligamiento.⁽¹⁹⁾

La fuerte correlación encontrada entre el rs3892097 y rs1065852 se explica por la corta distancia entre ambos (1747 bp), por lo que están físicamente relacionados y pueden tener un efecto estadístico conjunto en el cáncer de mama. Esto no siempre es de este modo, que puede suceder que dos loci que se encuentren a cortas distancias se encuentren en equilibrio de ligamiento, esto debido a ser variantes surgidas en tiempos remotos, lo que reduce probabilidad de recombinación.⁽²⁰⁾ La gran probabilidad de ligamiento, entonces, da lugar a tres escenarios posibles que: a) ambos SNPs tienen efecto causal sobre cáncer de mama; b) solo uno de los dos presenta un efecto casual, mientras que la asociación del otro se debe a ser siempre heredado con el primero; c) ninguno de los dos presenta un efecto casual, sino que están físicamente asociados a otro(s) SNPs que realmente tiene un efecto causal en el cáncer de mama (efecto confundente).⁽²¹⁾

CONCLUSIONES

Se concluye que los SNPs rs3892097 y rs1065852 presentan una alta probabilidad de ligamiento, convirtiéndose en candidatos para una asociación causal con el cáncer de mama. Por otro lado, el polimorfismo rs28371725 muestra una baja probabilidad, pudiendo presentar una asociación independiente con la enfermedad. La alta probabilidad de ligamiento entre rs3892097 y rs1065852, dada su proximidad física, descarta la asociación independiente a la enfermedad, pudiendo darse los tres escenarios mencionados.

REFERENCIAS

1. Avsar O. Identification of the effects of pathogenic genetic variations of human CYP2C9 and CYP2D6: an in silico approach. *Nucleosides Nucleotides Nucleic Acids* 2024;43:356-76. <https://doi.org/10.1080/15257770.2023.2262519>.
2. Hennig EE, Piątkowska M, Goryca K, Pośpiech E, Paziewska A, Karczmariski J, et al. Non-CYP2D6 variants selected by a GWAS improve the prediction of impaired tamoxifen metabolism in patients with breast cancer. *J Clin Med* 2019;8:1087. <https://doi.org/10.3390/jcm8081087>.
3. Sanchez-Spitman AB, Böhringer S, Dezentjé VO, Gelderblom H, Swen JJ, Guchelaar H-J. A genome-wide association study of endoxifen serum concentrations and adjuvant tamoxifen efficacy in early-stage breast cancer patients. *Clin Pharmacol Ther* 2024. <https://doi.org/10.1002/cpt.3255>.
4. Smith D, He B, Shi J, Zhu H-J, Wang X. Novel independent trans- and Cis-genetic variants associated with CYP2D6 expression and activity in human livers. *Drug Metab Dispos* 2024;52:143-52. <https://doi.org/10.1124/dmd.123.001548>.
5. Khor CC, Winter S, Sutiman N, Mürdter TE, Chen S, Lim JSL, et al. Cross-ancestry genome-wide association study defines the extended CYP2D6 locus as the principal genetic determinant of endoxifen plasma concentrations. *Clin Pharmacol Ther* 2023b;113:712-23. <https://doi.org/10.1002/cpt.2846>.
6. Goetz MP, Knox SK, Suman VJ, Rae JM, Safgren SL, Ames MM, et al. The impact of cytochrome P450 2D6 metabolism in women receiving adjuvant tamoxifen. *Breast Cancer Res Treat* 2007;101:113-21. <https://doi.org/10.1007/s10549-006-9428-0>.

7. Stingl JC (formerly, Parmar S, Huber-Wechselberger A, Kainz A, Renner W, Seeringer A, et al. Impact of CYP2D6*4 genotype on progression free survival in tamoxifen breast cancer treatment. *Curr Med Res Opin* 2010;26:2535-42. <https://doi.org/10.1185/03007995.2010.518304>.
8. Wegman P, Elingarami S, Carstensen J, Stål O, Nordenskjöld B, Wingren S. Genetic variants of CYP3A5, CYP2D6, SULT1A1, UGT2B15 and tamoxifen response in postmenopausal patients with breast cancer. *Breast Cancer Res* 2007;9. <https://doi.org/10.1186/bcr1640>.
9. Sukasem C, Chamnanphon M, Pechatanan K, Santon, Puangpetch, Chantratita W, et al. Association of CYP2D6 and CYP2C19 polymorphisms and disease-free survival of Thai post-menopausal breast cancer patients who received adjuvant tamoxifen. *Pharmgenomics Pers Med* 2013;37. <https://doi.org/10.2147/pgpm.s42330>.
10. Lim JSL, Chen XA, Singh O, Yap YS, Ng RCH, Wong NS, et al. Impact of CYP2D6, CYP3A5, CYP2C9 and CYP2C19 polymorphisms on tamoxifen pharmacokinetics in Asian breast cancer patients. *Br J Clin Pharmacol* 2011;71:737-50. <https://doi.org/10.1111/j.1365-2125.2011.03905.x>.
11. Schroth W, Antoniadou L, Fritz P, Schwab M, Muerdter T, Zanger UM, et al. Breast cancer treatment outcome with adjuvant tamoxifen relative to patient CYP2D6 and CYP2C19 genotypes. *J Clin Oncol* 2007;25:5187-93. <https://doi.org/10.1200/jco.2007.12.2705>.
12. Qanbari S. On the extent of linkage disequilibrium in the genome of farm animals. *Front Genet* 2020;10. <https://doi.org/10.3389/fgene.2019.01304>.
13. Fairley S, Lowy-Gallego E, Perry E, Flicek P. The International Genome Sample Resource (IGSR) collection of open human genomic variation resources. *Nucleic Acids Res* 2020;48:D941-7. <https://doi.org/10.1093/nar/gkz836>.
14. Byrska-Bishop M, Evani US, Zhao X, Basile AO, Abel HJ, Regier AA, et al. High-coverage whole-genome sequencing of the expanded 1000 Genomes Project cohort including 602 trios. *Cell* 2022;185:3426-3440.e19. <https://doi.org/10.1016/j.cell.2022.08.004>.
15. Danecek P, Auton A, Abecasis G, Albers CA, Banks E, DePristo MA, et al. The variant call format and VCFtools. *Bioinformatics* 2011;27:2156-8. <https://doi.org/10.1093/bioinformatics/btr330>.
16. Boullerne AI, Goudey B, Paganini J, Erlichster M, Gaitonde S, Feinstein DL. Validation of tag SNPs for multiple sclerosis HLA risk alleles across the 1000 genomes panel. *Hum Immunol* 2024;110790. <https://doi.org/10.1016/j.humimm.2024.110790>.
17. Wray NR. Allele frequencies and the measure of linkage disequilibrium: Impact on design and interpretation of association studies. *Twin Res Hum Genet* 2005;8:87-94. <https://doi.org/10.1375/twin.8.2.87>.
18. Habib I, Khan S, Mohammad T, Hussain A, Alajmi MF, Rehman T, et al. Impact of non-synonymous mutations on the structure and function of telomeric repeat binding factor 1. *J Biomol Struct Dyn* 2022;40:9053-66. <https://doi.org/10.1080/07391102.2021.1922313>.
19. Booker TR, Yeaman S, Whitlock MC. Variation in recombination rate affects detection of outliers in genome scans under neutrality. *bioRxiv* 2020. <https://doi.org/10.1101/2020.02.06.937813>.
20. Li R, Budowle B, Sun H, Ge J. Linkage and linkage disequilibrium among the markers in the forensic MPS panels. *J Forensic Sci* 2021;66:1637-46. <https://doi.org/10.1111/1556-4029.14724>.
21. Pingault J-B, Rijdsdijk F, Schoeler T, Choi SW, Selzam S, Krapohl E, et al. Genetic sensitivity analysis: Adjusting for genetic confounding in epidemiological associations. *PLoS Genet* 2021;17:e1009590. <https://doi.org/10.1371/journal.pgen.1009590>.

FINANCIACIÓN

Los autores no recibieron financiación para el desarrollo de la presente investigación.

CONFLICTO DE INTERESES

Los autores declaran que no existe conflicto de intereses.

CONTRIBUCIÓN DE AUTORÍA

Conceptualización: Sergio Vladimir Flores, Angel Roco-Videla, Raúl Aguilera-Eguía, Román Montaña.

Curación de datos: Sergio Vladimir Flores, Angel Roco-Videla.

Análisis formal: Sergio Vladimir Flores, Román Montaña.

Adquisición de fondos: Raúl Aguilera-Eguía.

Investigación: Sergio Vladimir Flores, Angel Roco-Videla, Raúl Aguilera-Eguía, Román Montaña.

Metodología: Sergio Vladimir Flores, Angel Roco-Videla.

Administración del proyecto: Sergio Vladimir Flores.

Recursos: Angel Roco-Videla.

Software: Sergio Vladimir Flores.

Supervisión: Román Montaña.

Validación: Sergio Vladimir Flores, Román Montaña.

Visualización: Sergio Vladimir Flores, Angel Roco-Videla.

Redacción - borrador original: Sergio Vladimir Flores, Angel Roco-Videla, Raúl Aguilera-Eguía, Román Montaña.

Redacción - revisión y edición: Angel Roco-Videla, Raúl Aguilera-Eguía.