

Análisis poblacional in silico del haplotipo CLOCK rs3749474T-rs4864548A y su relación con obesidad

Sergio Flores Carrasco PhD^a, Mariela Olguín-Barraza PhD^b, Angel Roco-Videla PhD^{c*}

^a Facultad de Ciencias de la Salud, Universidad Autónoma de Chile, Santiago, Chile

^b Facultad de Ciencias de la Salud, Programa de Magister en Ciencias Químico-Biológicas, Universidad Bernardo O'Higgins, Santiago, Chile

^c Escuela de Enfermería, Facultad de Salud y Ciencias Sociales, Universidad de las Américas, Santiago, Chile

*** Autor de correspondencia**
aroco@udla.cl

Citación

Flores Carrasco S, Olguín-Barraza M, Roco-Videla A. Análisis poblacional *in silico* del haplotipo CLOCK rs3749474T-rs4864548A y su relación con obesidad. *Medwave* 2023;23(9):e2735

DOI

10.5867/
medwave.2023.09.2735

Fecha de envío

Apr 7, 2023

Fecha de aceptación

Sep 13, 2023

Fecha de publicación

Oct 3, 2023

Palabras clave

Haplotypes, obesity, linkage disequilibrium, Period Circadian Proteins

Correspondencia a

Avenida Manuel Montt 948, Providencia, Santiago, Chile

Resumen

Se ha sugerido que el haplotipo rs3749474T/rs4864548A del gen CLOCK aumentaría el riesgo de obesidad, pero se desconoce el patrón de variabilidad poblacional de estos alelos y del haplotipo. El objetivo de este estudio es determinar el nivel de ligamiento entre los alelos de riesgo rs3749474T y rs4864548A a partir de la base de datos 1000Genomes para confirmar la existencia del haplotipo TA de los polimorfismos rs3749474-rs4864548 del gen CLOCK y su frecuencia cinco macro poblaciones. Se analizó el desequilibrio de ligamiento y las frecuencias haplotípicas para 2504 individuos, de 26 poblaciones, utilizando el estadístico r^2 y la prueba exacta de Fisher. Existe una alta frecuencia del haplotipo TA en Latinoamérica (44,8%), un alto desequilibrio de ligamiento ($r^2 = 0,92$) a nivel mundial entre esos alelos, una alta diferenciación entre macro poblaciones y una alta homogeneidad al interior de ellas. La evidencia presentada permite sugerir la realización de posteriores estudios de asociación entre este haplotipo y el nivel de riesgo de obesidad y sobrepeso en poblaciones latinoamericanas.

IDEAS CLAVE

- ◆ El haplotipo rs3749474T/rs4864548A del gen CLOCK aumentaría el riesgo de obesidad, pero se desconoce su patrón de variabilidad poblacional.
- ◆ Este estudio aporta antecedentes acerca de la frecuencia del haplotipo rs3749474T/rs4864548A del gen CLOCK en cinco macro poblaciones y 26 poblaciones.
- ◆ Existe una alta frecuencia del haplotipo TA en Latinoamérica, un alto desequilibrio de ligamiento, una alta diferenciación entre macro poblaciones y una alta homogeneidad al interior de ellas.
- ◆ Este estudio está limitado al análisis *in silico* de bases de datos, por lo que se sugiere confirmar la asociación del haplotipo TA con obesidad a través de estudios con datos primarios.

INTRODUCCIÓN

El gen CLOCK es uno de los genes que, a través de su proteína codificada, regula los ciclos circadianos, cuyos polimorfismos se han asociado al desarrollo de la obesidad, hipercolesterolemia, hiperglicemia entre otras condiciones [1]. Los SNPs rs3749474T y rs4864548A del gen CLOCK se han vinculado en forma independiente a un mayor índice de masa corporal, aumentando el riesgo de sobrepeso u obesidad hasta en 1,5 veces [2–6].

En el estudio realizado por Pino-Astorga *et al* se encontró el caso de un sujeto que presentaba ambos polimorfismos (rs3749474T y rs4864548A), cuyo perfil antropométrico lo sitúa como de alto riesgo debido a su sobrepeso y circunferencia de cintura, a pesar de tener hábitos alimenticios normales [7]. Lo anterior sugiere un efecto aditivo entre ambos alelos de riesgo, sugiriendo la existencia de del haplotipo TA para ellos.

El objetivo de esta investigación es determinar el nivel de ligamiento entre los alelos de riesgo rs3749474T y rs4864548A a partir de la base de datos 1000genome para confirmar la existencia del haplotipo TA de los polimorfismos rs3749474-rs4864548 del gen CLOCK, como un estudio preliminar a la realización de pruebas de campo confirmatorias en población latinoamericana.

MÉTODOS

MUESTRA

Se obtuvo los genotipos de los SNPs rs3749474 y rs4864548 del gen CLOCK de 2504 individuos, desde la base de datos del proyecto 1000Genomes en su fase 3 [8]. Para esto, se descargó el total de datos del cromosoma 4, para luego extraer los genotipos rs3749474 y rs4864548 para cada individuo, usando el software VCFtools. En dicho proyecto, todos los genomas fueron secuenciados a baja cobertura (4x)- Para la validación de datos, 24 de ellos fueron secuenciados con alta cobertura (50x). En este estudio se incluyeron las cinco macro poblaciones comprendidas en 1000Genomes: África, Este Asiático, Sur Asiático,

Europa y Latinoamérica, junto a las 26 poblaciones dentro de ellas (Tabla 1).

ANÁLISIS ESTADÍSTICO

La base 1000Genomes, en la actual fase de desarrollo contiene información “en fase” de los genotipos. Es decir, es posible reconstruir la distribución de alelos en cada cromosoma parental. Así, usando un algoritmo escrito en R, se identificaron los alelos de ambos SNPs para los 5008 cromosomas analizados, para luego estimar la frecuencia de los cuatro haplotipos según poblaciones y macro poblaciones.

Considerando los alelos ancestrales y derivados para ambos SNPs, así como sus respectivas frecuencias alélicas, se determinaron los haplotipos recombinantes y no-recombinantes, así como sus respectivas frecuencias poblacionales. Luego, se estimó la frecuencia de recombinación y el desequilibrio de ligamiento a través del estadístico r^2 . Las diferencias en las frecuencias de los haplotipos se analizaron usando la prueba exacta de Fisher.

RESULTADOS

Se obtuvo las siguientes frecuencias a nivel mundial, para los cuatro haplotipos de los SNPs rs3749474 y rs4864548: TA = 37,6%; CG = 62,2%; CA = 0,08% y TG = 0,08%. Considerando esas frecuencias, y que el haplotipo CG es ancestral según la información en 1000Genomes, resulta evidente que TA y CG son los haplotipos parentales en las actuales poblaciones. Así, la frecuencia de recombinación para estos dos *loci* es de 0,16 %. El desequilibrio de ligamiento fue alto ($r^2 = 0,92$).

La frecuencia de TA en las poblaciones se encuentra en la Tabla 1 y Figura 1. La menor frecuencia de genotipos portadores del haplotipo TA se encuentra en África, seguida por Europa, Latinoamérica, Sur Asiático y Este Asiático. Al comparar las frecuencias genotípicas entre todos los pares de poblaciones, se confirma la alta divergencia de África. Existen diferencias significativas ($p < 0,05$) en todas las comparaciones, excepto entre Latinoamérica y Sur Asiático ($p < 0,62$). Existe

Tabla 1. Información sobre las macro poblaciones y poblaciones analizadas en este estudio y la frecuencia del haplotipo TA en cada una.

Macro población (código)	Población	Código	N	% TA
África (AFR)	Caribe-Africanos en Barbados	ACB	96	16,7
	Ancestría africana en el sur oeste de EEUU	ASW	61	18,9
	Esan en Nigeria	ESN	99	21,2
	Gambianos del oeste	GWD	113	17,7
	Luhyas en Webuye, Kenia	LWK	99	13,1
	Mendés en Sierra Leona	MSL	85	14,7
	Yorubas en Ibadán, Nigeria	YRI	108	18,1
	Total		661	17,2
América latina (AMR)	Colombianos en Medellín	CLM	94	38,3
	Ancestría mexicana en Los Ángeles, EEUU	MXL	64	43,0
	Peruanos en Lima	PEL	85	58,8
	Puertorriqueños en Puerto Rico	PUR	104	40,4
	Total		347	44,8
Este Asiático (EAS)	Chinos Dai en Xishuangbanna, China	CDX	93	53,8
	Chinos Han en Beijing, China	CHB	103	61,7
	Chinos Han del sur	CHS	105	57,1
	Japoneses en Tokio	JPT	104	58,2
	Kinh en Ciudad Ho Chi Minh, Vietnam	KHV	99	54,0
	Total		504	57,0
Europa (EUR)	Utahéño con ancestría del norte y oeste europeo	CEU	99	40,9
	Fineses en Finlandia	FIN	99	35,7
	Británicos en Inglaterra y Escocia	GBR	91	32,8
	Ibéricos en España	IBS	107	30,4
	Toscanos en Italia	TSI	107	31,8
	Total		503	34,2
Sur Asiático (SAS)	Bengalíes en Bangladesh	BEB	99	45,9
	Gujarati en Houston, Texas	GIH	99	44,6
	Telugu en Reino Unido	ITU	91	41,3
	Punjabi en Lahore, Pakistán	PJL	107	44,8
	Sri Lanka Tamil en Reino Unido	STU	107	41,7
	Total		489	43,6

EEUU: Estados Unidos.
Fuente: elaboración propia.

alta homogeneidad dentro de las cinco macro poblaciones. Perú presenta una alta frecuencia, estadísticamente significativa, de TA en relación con las otras poblaciones de Latinoamérica ($p < 0,05$).

DISCUSIÓN

El alto nivel de ligamiento entre los alelos de riesgo rs3749474T y rs4864548A del gen CLOCK, aquí reportado, refuerza la observación del estudio de Pino-Astorga *et al* [7], pues ambos alelos de riesgo se heredan en conjunto en aproximadamente un 44,8% de los cromosomas en la población latinoamericana, mientras que en el restante 55,2% lo hace el haplotipo de no riesgo CG.

La alta divergencia de la macro población de África, respecto a las otras, podría deberse a la interacción de patrones dietarios y adaptación biológica [9], así como a cambios dietarios

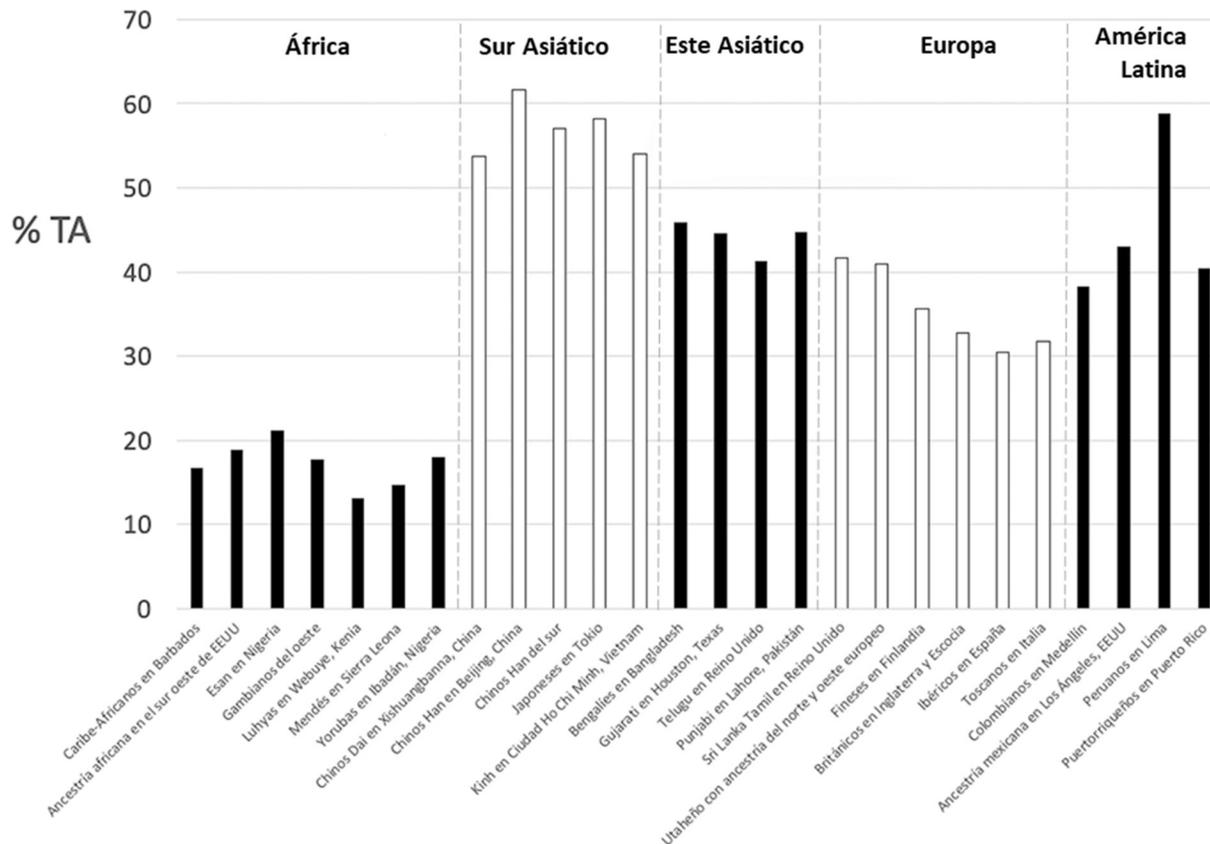
experimentados durante la transición alimentaria [10]. Las diferencias entre las restantes cuatro macro poblaciones podrían deberse a presiones selectivas específicas, durante el poblamiento de los diferentes continentes.

Llama la atención la alta homogeneidad entre poblaciones dentro de cada macro población, lo que refuerza nuestra hipótesis sobre la heterogeneidad entre macro poblaciones. La alta divergencia observada en la población de Lima, que exhibió una alta frecuencia del haplotipo TA, requiere de posteriores análisis. Una posibilidad de explicación a este hecho es la mayor acumulación de alelos de origen asiático, considerando la historia de poblamiento y mestizaje ocurrida en Perú.

Una limitación del presente trabajo es la existencia de particularidades poblacionales que aquí no son abarcadas. Ejemplo de ello es la existencia de diferentes ancestrías dentro de las macro poblaciones, como es el caso de poblaciones indígenas en Latinoamérica. Esta necesidad se hace evidente al observar la

Figura 1: Frecuencia del haplotipo TA en cada una de las poblaciones y macro poblaciones analizadas en este estudio.

El tamaño muestral para un margen error de 5% y un intervalo de confianza de 95% es de $n = 385$. Solo una macro población, la latinoamericana, presenta un tamaño muestral bajo ese valor, correspondiendo a 347 participantes. Para ese tamaño de muestra, el margen de error es de 5,15%, es decir muy próximo al 5%.



Fuente: elaboración propia.

alta frecuencia del haplotipo TA en población de Lima, Perú. Esta presenta una mayor ancestría nativo americana, en relación con las otras poblaciones latinoamericanas aquí estudiadas. No obstante esta limitación, la evidencia recolectada permite establecer la existencia del haplotipo TA rs3749474-rs4864548 con una relativa alta frecuencia y desequilibrio de ligamiento, siendo un marcador candidato para un estudio de riesgo de sobrepeso y obesidad en población latinoamericana.

CONCLUSIONES

El alto nivel de ligamiento entre los alelos de riesgo rs3749474T y rs4864548A, junto a su relativa alta frecuencia en población latinoamericana, permiten confirmar la existencia del haplotipo TA para estos polimorfismos. Ello lo convierte en un candidato de investigación a nivel población, para establecer su relación con el desarrollo de la obesidad y sobrepeso.

Notas

Autoría

SVF, MOB, ARV: concepción y diseño del trabajo, redacción del manuscrito, revisión crítica del manuscrito, aprobación de su versión final. SVF, ARV: recolección de datos. SVF: análisis de datos e interpretación de resultados.

Conflictos de intereses

Los autores declaran no tener conflicto de intereses.

Financiamiento

Los autores declaran que no hubo fuentes de financiamiento externo.

Origen y arbitraje

No solicitado. Con revisión externa por tres pares revisores, a doble ciego.

Idioma del envío

Español.

Referencias

1. Turek FW, Joshu C, Kohsaka A, Lin E, Ivanova G, McDearmon E, et al. Obesity and metabolic syndrome in circadian Clock mutant mice. *Science*. 2005;308: 1043–5. <https://doi.org/10.1126/science.1108750>
2. Sookoian S, Gemma C, Gianotti TF, Burgueño A, Castaño G, Pirola CJ. Genetic variants of Clock transcription factor are associated with individual susceptibility to obesity. *Am J Clin Nutr*. 2008;87: 1606–15. <https://doi.org/10.1093/ajcn/87.6.1606> <https://doi.org/10.1093/ajcn/87.6.1606>
3. Scott EM, Carter AM, Grant PJ. Association between polymorphisms in the Clock gene, obesity and the metabolic syndrome in man. *Int J Obes*. 2008;32: 658–662. <https://doi.org/10.1038/sj.ijo.0803778> <https://doi.org/10.1038/sj.ijo.0803778>
4. Garaulet M, Lee Y-C, Shen J, Parnell LD, Arnett DK, Tsai MY, et al. CLOCK genetic variation and metabolic syndrome risk: modulation by monounsaturated fatty acids. *Am J Clin Nutr*. 2009;90: 1466–75. <https://doi.org/10.3945/ajcn.2009.27536> <https://doi.org/10.3945/ajcn.2009.27536>
5. Garaulet M, Lee Y-C, Shen J, Parnell LD, Arnett DK, Tsai MY, et al. Genetic variants in human CLOCK associate with total energy intake and cytokine sleep factors in overweight subjects (GOLDN population). *Eur J Hum Genet*. 2010;18: 364–9. <https://doi.org/10.1038/ejhg.2009.176> <https://doi.org/10.1038/ejhg.2009.176>
6. Loria-Kohen V, Espinosa-Salinas I, Marcos-Pasero H, Lourenço-Nogueira T, Herranz J, Molina S, et al. Polymorphism in the CLOCK gene may influence the effect of fat intake reduction on weight loss. *Nutrition*. 2016;32: 453–60. <https://doi.org/10.1016/j.nut.2015.10.013>
7. Pino-astorga C, Roco-Videla Agerardo, Valladares Vega M. Potential Joint Effect of the SNP rs3749474T and SNPrs4864548A Alleles of the CLOCK Gene on Nutritional Status and Chronotype in Young University Students. *sun*. 2022;37: 606–618. <https://rcientificas.uninorte.edu.co/index.php/salud/issue/view/587> <https://doi.org/10.14482/sun.37.1.613.2>
8. Auton A, Brooks LD, Durbin RM, Garrison EP, Kang HM, Korbel JO, et al. A global reference for human genetic variation. *Nature*. 2015;526: 68–74. <https://doi.org/10.1038/nature15393>
9. Luca F, Perry GH, Di Rienzo A. Evolutionary adaptations to dietary changes. *Annu Rev Nutr*. 2010;30: 291–314. <https://doi.org/10.1146/annurev-nutr-080508-141048>
10. Popkin BM, Ng SW. The nutrition transition to a stage of high obesity and noncommunicable disease prevalence dominated by ultra-processed foods is not inevitable. *Obes Rev*. 2022;23. <https://doi.org/10.1111/obr.13366>

In silico population analysis of the CLOCK rs3749474T-rs4864548A haplotype and its relationship with obesity

Abstract

It has been suggested that the rs3749474T/rs4864548A haplotype of the CLOCK gene increases the risk of obesity, but the population variability of these alleles and the haplotype is unknown. This research aims to determine the linkage between the rs3749474T and rs4864548A alleles from the database of 1000Genomes to confirm the existence of the TA haplotype polymorphisms of these alleles and their frequency in five macro populations. Linkage disequilibrium and haplotype frequencies for 2504 individuals from 26 populations were analyzed using the r^2 statistic and Fisher's exact test. There is a high frequency of the TA haplotype in Latin America (44.8%), a high linkage disequilibrium ($r^2= 0.92$) worldwide between these alleles, a high differentiation between macro populations, and a high homogeneity. The evidence warrants further studies on the association between this haplotype and the risk of obesity and overweight in Latin American populations.



This work is licensed under a Creative Commons Attribution 4.0 International License.