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

Sergio Flores Carrasco[a], Mariela Olguín-Barraza[b], Angel Roco-Videla[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

* Corresponding author
aroco@udla.cl

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 (r2= 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.

MAIN MESSAGES

♦ The rs3749474T/rs4864548A haplotype of the CLOCK gene increases the risk of obesity, but population variability is unknown.
♦ This study provides background on the frequency of the rs374947474T/rs4864548A haplotype of the CLOCK gene in 26 populations among five macropopulations.
♦ There is a high frequency of the TA haplotype in Latin America, high linkage disequilibrium, high differentiation between macro populations, and high homogeneity within populations.
♦ This study is limited to in silico analysis of databases and warrants further primary studies to confirm the association of the TA haplotype with obesity.

INTRODUCTION

The CLOCK gene regulates circadian cycles through its encoded protein. Its polymorphisms have been associated with obesity, hypercholesterolemia, and hyperglycemia, among other conditions [1]. The single-nucleotide polymorphisms (SNPs) rs3749474T and rs4864548A of the CLOCK gene have been independently linked to a higher body mass index, increasing the risk of overweight and obesity by up to 1.5-fold [2–6].

In the study by Pino-Astorga et al., a subject with both polymorphisms (rs3749474T and rs4864548A) was overweight and had a high waist circumference despite having normal eating habits [7]. This implies an additive effect between both risk alleles, suggesting the existence of the TA haplotype.

This research is a preliminary study aiming to determine the linkage between the rs3749474T and rs4864548A risk alleles from the 1000genome database to confirm the existence of the TA haplotype of the rs3749474-rs4864548 polymorphisms of the CLOCK gene.

METHODS

SAMPLE

The genotypes of SNPs rs3749474 and rs4864548 of the CLOCK gene were obtained from 2504 individuals from the 1000Genomes project database in phase 3 [8]. The total chromosome 4 data was downloaded, and then the rs3749474 and rs4864548 genotypes were extracted for each individual using VCFtools software. In that project, all genomes were sequenced at low coverage (4x). For data validation, 24 were sequenced at high coverage (50x). The five macro populations included in 1000Genomes were included in this study: Africa, East Asia, South Asia, Europe, and Latin America, together with the 26 populations within them (Table 1).

The sample size for a margin of error of 5% and a confidence interval of 95% was n = 385. Only Latin American was the macropopulation with a sample size below this value, corresponding to 347 participants. For this sample size, the margin of error is 5.15%.

STATISTICAL ANALYSIS

At the current stage of development, the 1000Genomes database contains "in-phase" information on genotypes. Therefore, it is possible to reconstruct the allele distribution on each parental chromosome. Using an algorithm written in R, we identified the alleles of both SNPs for the 5008 chromosomes analyzed and then estimated the frequency of the four haplotypes according to populations and macro populations.

Considering the ancestral and derived alleles for both SNPs and their respective allele frequencies, the recombinant and non-recombinant haplotypes and their respective population frequencies were determined. Then, recombination frequency and linkage disequilibrium were estimated through the r2 statistic. Differences in haplotype frequencies were analyzed using Fisher's exact test.

RESULTS

The following frequencies were obtained worldwide for the four haplotypes of SNPs of rs3749474 and rs4864548: TA = 37.6%, CG = 62.2%, CA = 0.08%, and TG = 0.08%. Considering these frequencies and that the CG haplotype is ancestral according to the information in 1000Genomes, it is evident that TA and CG are the parental haplotypes in the present populations. Thus, the recombination frequency for these two loci is 0.16 %. Linkage disequilibrium was high (r2= 0.92).

The frequency of TA in the populations is given in Table 1 and Figure 1. The lowest frequency of TA haplotype genotypes is found in Africa, followed by Europe, Latin America, South Asia, and East Asia. Comparing genotypic frequencies between all pairs of populations confirms the high divergence from Africa. There are significant differences (p < 0.05) in all comparisons except between Latin America and South Asia (p < 0.62). However, there is high homogeneity within the five macro populations. Peru has a statistically significant higher

Table 1. Information on the macro populations and populations analyzed in this study and the frequency of the TA haplotype in each of them.

Macro population (code)	Population	Code	N	% TA
Africa (AFR)	Caribbean-Africans in Barbados	ACB	96	16.7
	African ancestry in the southwestern USA.	ASW	61	18.9
	Esan in Nigeria	ESN	99	21.2
	Western Gambians	GWD	113	17.7
	Luhyas in Webuye, Kenya	LWK	99	13.1
	Mendés in Sierra Leone	MSL	85	14.7
	Yorubas in Ibadan, Nigeria	YRI	108	18.1
		Total	**661**	**17.2**
America (AMR)	Colombians in Medellín	CLM	94	38.3
	Mexican ancestry in Los Angeles, USA	MXL	64	43.0
	Peruvians in Lima	PEL	85	58.8
	Puerto Ricans in Puerto Rico	PUR	104	40.4
		Total	**347**	**44.8**
East Asia (EAS)	Dai Chinese in Xishuangbanna, China	CDX	93	53.8
	Han Chinese in Beijing, China	CHB	103	61.7
	Southern Han Chinese	CHS	105	57.1
	Japanese in Tokyo	JPT	104	58.2
	Kinh in Ho Chi Minh City, Vietnam	KHV	99	54.0
		Total	**504**	**57.0**
Europe (EUR)	Utaheño with ancestry from Northern and Western Europe	CEU	99	40.9
	Finns in Finland	FIN	99	35.7
	British in England and Scotland	GBR	91	32.8
	Iberians in Spain	IBS	107	30.4
	Tuscans in Italy	TSI	107	31.8
		Total	**503**	**34.2**
South Asia (SAS)	Bengalis in Bangladesh	BEB	99	45.9
	Gujarati in Houston, Texas	GIH	99	44.6
	Telugu in the United Kingdom	ITU	91	41.3
	Punjabi in Lahore, Pakistan	PJL	107	44.8
	Sri Lankan Tamil in the United Kingdom	STU	107	41.7
		Total	**489**	**43.6**

USA: United States.
Source: Prepared by the authors of this study.

frequency of TA relative to the other Latin American populations ($p < 0.05$).

DISCUSSION

The high level of linkage between the rs3749474T and rs4864548A risk alleles of the CLOCK gene reported here reinforces the observation of the study by Pino-Astorga et al. [7], as both risk alleles are jointly inherited in approximately 44.8% of the chromosomes in the Latin American population. In comparison, in the remaining 55.2%, the non-risk CG haplotype is inherited.

The high divergence of the African macro population could be due to the interaction of dietary patterns and biological adaptation [9] and dietary changes experienced during the food transition [10]. The differences among the remaining four macro populations could be due to specific selective pressures during the settlement of the different continents.

The high homogeneity between populations within each macro population is striking, which reinforces our hypothesis of heterogeneity between macro populations. The high divergence observed in the Lima population, which exhibited a high frequency of the TA haplotype, requires further analysis. One possible explanation for this is the greater accumulation of alleles of Asian origin, considering the history of settlement and miscegenation in Peru.

One limitation of the present work is the existence of population particularities that are not covered here. An example of this is the existence of different ancestries within macro populations, as is the case of indigenous populations in Latin America. This need becomes evident when observing the high frequency of the TA haplotype in Lima, Peru. This country presents a higher Native American ancestry than the other

Figure 1. Frequency of the TA haplotype in each population and macro population analyzed in this study.
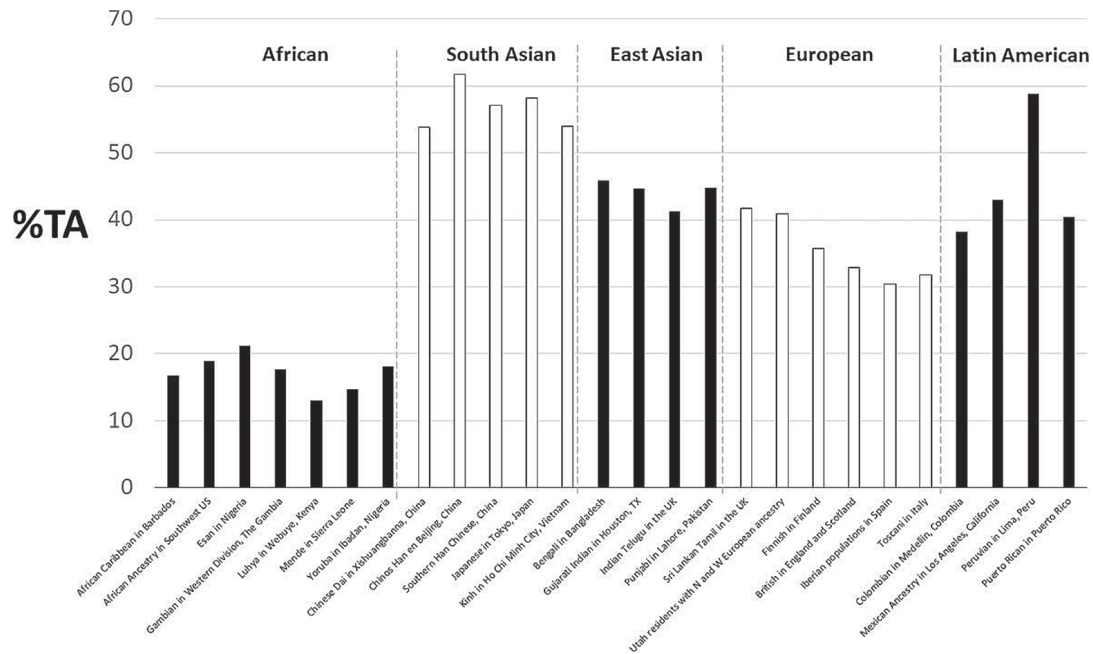
Source: Prepared by the authors of this study.

Latin American populations studied here. Despite this limitation, the evidence collected establishes the existence of the TA haplotype rs3749474-rs4864548 with a relatively high frequency and linkage disequilibrium, a candidate marker for a risk study of overweight and obesity in the Latin American population.

CONCLUSIONS

The high level of linkage between the rs3749474T and rs4864548A risk alleles, together with their relatively high frequency in the Latin American population, allow us to confirm the existence of the TA haplotype for these polymorphisms. This finding warrants further research at a population level to establish its relationship with the development of obesity and overweight.

Notes

Contributor roles

SVF, MOB, ARV: conception and design of the work, drafting of the manuscript, critical revision of the manuscript, approval of its final version. SVF, ARV: data collection. SVF: data analysis and interpretation of results.

Competing interests

The authors declare that they have no conflicts of interest.

Funding

The authors state that there were no external sources of financing.

Provenance and peer review

Unsolicited. With external review by three peer reviewers, double-blind.

Language of submission

Spanish.

References

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 A gerardo, Valladares Vega M. Potential Joint Effect of the SNP rs3749474T and SNP rs4864548A 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

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

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.