



**UNIVERSIDAD NACIONAL AUTÓNOMA DE MÉXICO**

**POSGRADO EN CIENCIAS BIOLÓGICAS**

FACULTAD DE MEDICINA

BIOLOGÍA EXPERIMENTAL

**IMPACTO DE LAS VARIANTES CON EVIDENCIA DE SELECCIÓN NATURAL  
POSITIVA EN EL GEN *PPARG* SOBRE LOS PERFILES DE EXPRESIÓN EN TEJIDO**

**ADIPOSO**

**TESIS**

QUE PARA OPTAR POR EL GRADO DE:

**MAESTRO EN CIENCIAS BIOLÓGICAS**

PRESENTA:

**QFB. LÓPEZ ESCOBAR TOMAS ALEJANDRO**

TUTOR PRINCIPAL DE TESIS:

**DR. HUMBERTO GARCÍA ORTIZ**

FACULTAD DE MEDICINA, UNAM

COMITÉ TUTOR:

**DRA. LORENA SOFÍA OROZCO OROZCO**

FACULTAD DE MEDICINA, UNAM

**DRA. LORENA AGUILAR ARNAL**

INSTITUTO DE INVESTIGACIONES BIOMÉDICAS, UNAM

**Ciudad Universitaria, CD. MX., 2022**



Universidad Nacional  
Autónoma de México

Dirección General de Bibliotecas de la UNAM

**Biblioteca Central**



**UNAM – Dirección General de Bibliotecas**  
**Tesis Digitales**  
**Restricciones de uso**

**DERECHOS RESERVADOS ©**  
**PROHIBIDA SU REPRODUCCIÓN TOTAL O PARCIAL**

Todo el material contenido en esta tesis esta protegido por la Ley Federal del Derecho de Autor (LFDA) de los Estados Unidos Mexicanos (México).

El uso de imágenes, fragmentos de videos, y demás material que sea objeto de protección de los derechos de autor, será exclusivamente para fines educativos e informativos y deberá citar la fuente donde la obtuvo mencionando el autor o autores. Cualquier uso distinto como el lucro, reproducción, edición o modificación, será perseguido y sancionado por el respectivo titular de los Derechos de Autor.





**UNIVERSIDAD NACIONAL AUTÓNOMA DE MÉXICO**

**POSGRADO EN CIENCIAS BIOLÓGICAS**

FACULTAD DE MEDICINA

BIOLOGÍA EXPERIMENTAL

**IMPACTO DE LAS VARIANTES CON EVIDENCIA DE SELECCIÓN NATURAL  
POSITIVA EN EL GEN *PPARG* SOBRE LOS PERFILES DE EXPRESIÓN EN TEJIDO**

**ADIPOSO**

**TESIS**

QUE PARA OPTAR POR EL GRADO DE:

**MAESTRO EN CIENCIAS BIOLÓGICAS**

PRESENTA:

**QFB. LÓPEZ ESCOBAR TOMAS ALEJANDRO**

TUTOR PRINCIPAL DE TESIS:

**DR. HUMBERTO GARCÍA ORTIZ**

FACULTAD DE MEDICINA, UNAM

COMITÉ TUTOR:

**DRA. LORENA SOFÍA OROZCO OROZCO**

FACULTAD DE MEDICINA, UNAM

**DRA. LORENA AGUILAR ARNAL**

INSTITUTO DE INVESTIGACIONES BIOMÉDICAS, UNAM

**Ciudad Universitaria, CD. MX., 2022**

COORDINACIÓN DEL POSGRADO EN CIENCIAS BIOLÓGICAS

ENTIDAD FACULTAD DE MEDICINA

OFICIO CPCB/1194/2021

ASUNTO: Oficio de Jurado

M. en C. Ivonne Ramírez Wence  
Directora General de Administración Escolar, UNAM  
Presente

Me permito informar a usted que en la reunión ordinaria del Comité Académico del Posgrado en Ciencias Biológicas, celebrada el día **9 de agosto de 2021** se aprobó el siguiente jurado para el examen de grado de **MAESTRO EN CIENCIAS BIOLÓGICAS** en el campo de conocimiento de **Biología Experimental** del estudiante **LÓPEZ ESCOBAR TOMAS ALEJANDRO** con número de cuenta **520010239** con la tesis titulada **“IMPACTO DE LAS VARIANTES CON EVIDENCIA DE SELECCIÓN POSITIVA EN EL GEN PPARG SOBRE LOS PERFILES DE EXPRESIÓN DE TEJIDO ADIPOSO”**, realizada bajo la dirección del **DR. HUMBERTO GARCÍA ORTIZ**, quedando integrado de la siguiente manera:

Presidente: DRA. MARCIA HIRIART URDANIVIA  
Vocal: DRA. ANGÉLICA GRACIELA MARTÍNEZ HERNÁNDEZ  
Vocal: DRA. SANDRA ROMERO HIDALGO  
Vocal: DR. EMILIO JOAQUÍN CORDOVA ALARCÓN  
Secretario: DRA. LORENA SOFÍA OROZCO OROZCO

Sin otro particular, me es grato enviarle un cordial saludo.

**ATENTAMENTE**  
**“POR MI RAZA HABLARÁ EL ESPÍRITU”**  
Ciudad Universitaria, Cd. Mx., a 07 de diciembre de 2021

**COORDINADOR DEL PROGRAMA**



**DR. ADOLFO GERARDO NAVARRO SIGÜENZA**



**COORDINACIÓN DEL POSGRADO EN CIENCIAS BIOLÓGICAS**

Unidad de Posgrado, Edificio D, 1º Piso. Circuito de Posgrados, Ciudad Universitaria  
Alcaldía Coyoacán. C. P. 04510 CDMX Tel. (+5255)5623 7002 <http://pcbiol.posgrado.unam.mx/>

## **Agradecimientos institucionales**

Al posgrado en Ciencias Biológicas de la Universidad Nacional Autónoma de México por darme las herramientas necesarias para poder realizar mis estudios de maestría.

Al Consejo Nacional de Ciencia y Tecnología por la beca otorgada sin la cual no habría podido realizar mis estudios de maestría y esta tesis. CVU: 1000674

A mi tutor, el Dr. Humberto García Ortiz, por todo su apoyo a lo largo del proceso de elaboración, sus enseñanzas y consejos a lo largo de mi formación como Maestro en Ciencias.

A los miembros del Comité Tutorial, la Dra. Lorena Sofía Orozco Orozco y la Dra. Lorena Aguilar Arnal, por su tiempo, consejos y comentarios que sin duda contribuyeron constructivamente al desarrollo de esta tesis.

## **Agradecimientos a título personal**

A mis padres, Reyna de Jesús Escobar Guzmán y Cruz Miguel López Flores, por brindarme amor, apoyo incondicional, guía, confianza, y un sinfín de cualidades sin las cuales no sería la persona que soy hoy en día. Mis frutos producto de sus grandes esfuerzos.

A mi hermano, Sergio Iván López Escobar, que sin sus palabras de aliento y ánimo no podría haber llevado a cabo este logro.

A Erika Vanessa Molina Murillo, por su gran apoyo, paciencia y amor a lo largo de todos estos años para la realización de este y todos mis proyectos, nuestros logros.

A la Dra. Ana Olivia Cañas Urbina, mi tutora de licenciatura y gran amiga, de quien he aprendido mucho de sus lecciones no solo académicas, también de vida.

A mi tutor, el Dr. Humberto García Ortiz, por brindarme confianza para formar parte de su grupo de trabajo y llevar a cabo este proyecto.

A mis amigos, David y Diana y a todos mis compañeros de laboratorio que, mientras las condiciones sanitarias lo permitieron, fueron una gran compañía que hicieron más amena mi estancia en el laboratorio a lo largo de mis estudios de maestría.

# ÍNDICE

## ÍNDICE DE FIGURAS

## ÍNDICE DE TABLAS

## ÍNDICE DE ECUACIONES

I.	RESUMEN .....	1
1.1	ABSTRACT .....	3
II.	INTRODUCCIÓN.....	5
2.1	SELECCIÓN NATURAL.....	5
2.2	VARIACIÓN GENÉTICA .....	8
2.3	DESEQUILIBRIO DE LIGAMIENTO .....	10
2.4	SELECCIÓN NATURAL EN POBLACIONES HUMANAS .....	11
2.5	GEN <i>PPARG</i> .....	14
III.	JUSTIFICACIÓN.....	18
IV.	OBJETIVOS .....	20
4.1	OBJETIVO GENERAL.....	20
4.2	OBJETIVOS ESPECÍFICOS .....	20
V.	METODOLOGÍA .....	21
5.1	POBLACIÓN DE ESTUDIO .....	21
5.2	IDENTIFICACIÓN DE LA ESTRUCTURA DE LOS HAPLOTIPOS BAJO SELECCIÓN .....	22
5.3	ANOTACIÓN FUNCIONAL <i>IN SILICO</i> DE LAS VARIANTES QUE COMPONEN EL HAPLOTIPO BAJO SELECCIÓN EN <i>PPARG</i> .....	22
5.4	DETERMINACIÓN DEL COEFICIENTE DE SELECCIÓN DEL HAPLOTIPO DE <i>PPARG</i> .....	23
5.5	ESTIMACIÓN DEL TIEMPO DE SELECCIÓN DEL HAPLOTIPO DE <i>PPARG</i> .....	24
5.6	SELECCIÓN DE MUESTRAS DE TEJIDO ADIPOSEO.....	25
VI.	RESULTADOS.....	28
6.1	ESTRUCTURA Y ANÁLISIS FUNCIONAL <i>IN SILICO</i> DE LOS HAPLOTIPOS.....	28
6.2	COEFICIENTE DE SELECCIÓN .....	32
6.3	TIEMPO AL ANCESTRO COMÚN MÁS RECIENTE (TMRCA).....	33
6.4	ANÁLISIS DIFERENCIAL GENÓMICO.....	36
VII.	DISCUSIÓN.....	43
VIII.	CONCLUSIONES .....	48
IX.	REFERENCIAS BIBLIOGRÁFICAS .....	49
X.	ANEXOS .....	62
10.1	GRAFICAS DE DESEQUILIBRIO DE LIGAMIENTO EN EL GEN <i>PPARG</i> EN LAS POBLACIONES DE LA FASE 3 DEL PROYECTO DE LOS 1000 GENOMAS .....	62



## Índice de figuras

FIGURA 1.- LA FUERZA DE LA SELECCIÓN NATURAL INFLUYE EN LA TASA DE CAMBIO EN LAS FRECUENCIAS GENOTÍPICAS Y ALÉLICAS. ....	6
FIGURA 2.- TIPOS DE SELECCIÓN NATURAL EN RASGOS CUANTITATIVOS. ....	7
FIGURA 3.- DISTRIBUCIÓN HIPOTÉTICA DE LOS EFECTOS DE LAS MUTACIONES EN EL FENOTIPO QUE FINALMENTE IMPACTAN EN EL FITNESS DEL GENOTIPO. ....	9
FIGURA 4.- EJEMPLOS DE ADAPTACIONES LOCALES HUMANAS. ....	12
FIGURA 5.- SEÑALES COMPARTIDAS DE SELECCIÓN A TRAVÉS DE POBLACIONES INDÍGENAS DE DIFERENTES REGIONES DE MÉXICO. ....	13
FIGURA 7.- ESTRUCTURA DEL HAPLOTIPO BAJO SELECCIÓN DE PPARG EN POBLACIÓN INDÍGENA Y POBLACIONES DEL PROYECTO 1000 GENOMAS. ....	29
FIGURA 8.- INTERACCIONES PROMOTOR-ENHANCER DEL GEN PPARG EN DATOS DE HIC DE TEJIDO ADIPOSO Y RNA-SEQ DE 7 LÍNEAS CELULARES DEL PROYECTO ENCODE. ....	31
FIGURA 9.- TIEMPO AL ANCESTRO COMÚN MÁS RECIENTE (TMRCA) EN GENERACIONES DE LOS MODELOS IMPLEMENTADOS CON EL SOFTWARE STARTMRCA. ....	34
FIGURA 10.- MEDIAS DEL TMRCA EN GENERACIONES DE LAS VARIANTES CON VALORES ALTOS DE SELECCIÓN. ....	35
FIGURA 11.- NIVELES DE EXPRESIÓN DEL GEN PPARG ENTRE PORTADORES DEL HAPLOTIPO SILVESTRE EN ESTADO HOMOCIGOTO (HS, N = 3) Y HETEROCIGOTOS QUE PORTAN UNA COPIA DEL HAPLOTIPO BAJO SELECCIÓN (HE, N = 9). ....	37
FIGURA 12.- MAPA DE CALOR DE LOS 281 TRANSCRITOS DIFERENCIALMENTE EXPRESADOS ENTRE HS Y HE. ....	38
FIGURA 13.- RED DE INTERACCIONES DE LOS GENES EXPRESADOS DIFERENCIALMENTE DE MANERA SIGNIFICATIVA (N = 281 GENES, FC >  1 , P < 0.05) ENTRE LOS INDIVIDUOS HOMOCIGOTOS SILVESTRES Y HETEROCIGOTOS. ....	40
FIGURA 14.- ANÁLISIS DE ENRIQUECIMIENTO DE LOS PROCESOS BIOLÓGICOS EN LOS QUE PARTICIPAN LOS GENES EXPRESADOS DIFERENCIALMENTE DE MANERA SIGNIFICATIVA (N = 281 GENES, FC >  1 , P < 0.05) ENTRE LOS INDIVIDUOS HS Y HE. ....	41
FIGURA 15.- ANÁLISIS DE ENRIQUECIMIENTO DE LAS FUNCIONES MOLECULARES EN LAS QUE PARTICIPAN LOS GENES EXPRESADOS DIFERENCIALMENTE DE MANERA SIGNIFICATIVA ENTRE LOS INDIVIDUOS HS Y HE (N = 281 GENES, FC >  1 , P < 0.05). ....	41
FIGURA 16.- ANÁLISIS DE ENRIQUECIMIENTO DE LAS RUTAS METABÓLICAS EN LAS QUE PARTICIPAN LOS GENES EXPRESADOS DIFERENCIALMENTE DE MANERA SIGNIFICATIVA (N = 281 GENES, FC >  1 , P < 0.05) ENTRE LOS INDIVIDUOS HS Y HE. ....	42

## Índice de tablas

TABLA 1 VARIANTES CON EVIDENCIA DE SELECCIÓN EN EL GEN PPARG.....	14
TABLA 2 CARACTERÍSTICAS BIOQUÍMICAS Y ANTROPOMÉTRICAS DE LOS 12 INDIVIDUOS MESTIZOS CON OBESIDAD MÓRBIDA SELECCIONADOS PARA EL ANÁLISIS DIFERENCIAL GENÓMICO. ....	27
TABLA 3.- CONSECUENCIA DE LAS VARIANTES DEL HAPLOTIPO BAJO SELECCIÓN (N = 35). ....	30
TABLA 4.- COEFICIENTE DE SELECCIÓN DE LA VARIANTE RS9833097 CALCULADO A PARTIR DE LA FRECUENCIA ALÉLICA DE LA VARIANTE RS6782475 CON LA QUE SE ENCUENTRA EN DESEQUILIBRIO DE LIGAMIENTO. ....	33
TABLA 5.- COMPARACIÓN DE LAS MEDIAS OBTENIDAS POR EL BOOTSTRAPPING (N=10) DEL TMRCA DE LAS 4 VARIANTES CON VALORES MÁS ALTOS DE SELECCIÓN. ....	36
TABLA 6.- TOP 20 DE LOS GENES CON LOS VALORES MÁS SIGNIFICATIVOS ( $FC >  1 $ , $P < 0.05$ ). ....	39

## Índice de ecuaciones

ECUACIÓN 1.- MODELO SIMPLE PARA ESTIMAR EL COEFICIENTE DE SELECCIÓN.....	24
--	----

## I. Resumen

La selección natural es un proceso que ha participado en la historia evolutiva de la mayoría de las especies, incluyendo al humano. Debido a ello, se ha realizado la búsqueda de evidencia de este proceso evolutivo a lo largo de todo el genoma humano en distintas poblaciones. Como resultado, en diferentes genes se han identificado variantes con señales de procesos de selección reciente y que podrían estar relacionadas a procesos de adaptación local al ambiente en el que se asentaron los diferentes grupos humanos alrededor del mundo. Un ejemplo de estos estudios es el realizado previamente por nuestro grupo de trabajo en el que se identificaron genes con señales de selección positiva en el genoma de indígenas de la cohorte MAIS (*Metabolic Analysis in an Indigenous Sample*) provenientes de 4 regiones geográficas de México. Uno de los genes con variantes con señales de selección compartidas entre estos individuos fue *PPARG* (rs9833097, rs7626560, rs6782475 y rs13090265), el cual es relevante biomédicamente ya que variantes en este gen han sido asociadas a padecimientos metabólicos como la diabetes y obesidad. Debido a lo anterior, el objetivo del presente trabajo fue caracterizar el proceso de selección que ocurrió en las variantes del gen *PPARG* en población indígena y determinar si estos alelos se asocian con perfiles transcripcionales específicos en tejido adiposo. Nuestros resultados revelaron un haplotipo específico en este gen para la población indígena de México que consta de 35 variantes ubicadas principalmente en regiones intrónicas y algunas de ellas en regiones reguladoras y que, posiblemente se seleccionó hace aproximadamente 15 mil años. Esta cantidad de años coincide con la que se estima en que ocurrió la migración de humanos modernos al continente americano. Dicho evento ocurrió en condiciones ambientales poco favorables como climas extremadamente fríos, poca disponibilidad de alimentos y la necesidad de recorrer grandes distancias a pie. Tomando en cuenta que este haplotipo se encuentra en *PPARG*, el cual regula procesos importantes relacionados al metabolismo energético como la adipogénesis, nuestros resultados sugieren que este evento de migración pudo ser la presión selectiva que contribuyó a la selección de estas variantes. Además, mediante análisis de expresión diferencial en tejido adiposo, identificamos 281

transcritos expresados diferencialmente en los individuos portadores del haplotipo bajo selección. Algunos de estos transcritos son importantes para procesos biológicos relacionados al metabolismo de lípidos. Dado que identificamos que algunas de las variantes que integran el haplotipo bajo selección sobre *PPARG* se encuentran en regiones reguladoras del gen, aunado a los resultados del análisis diferencial genómico, inferimos que el haplotipo bajo selección podría estar causando cambios en la regulación de la expresión del gen lo que a su vez modifica la expresión de los genes regulados por *PPARG*. Dando como resultado una firma de expresión que podría resultar en un manejo más eficaz en la energía del organismo de los primeros pobladores de América.

## 1.1 Abstract

Natural selection processes has been participated in evolutionary history of several species, including humans. Due this, it has been looked for evidence of this evolutionary process along whole human genome through several populations. In consequence, it has been identified several genes containing variants with signals of recent selection processes and that could be related to local adaptation processes to the specific environment in which these groups settled. An example of this kind of studies is the one previously done for our work group in which it was identified several genes with signals of positive selection through genomes of indigenous from 4 geographic regions of Mexico. One of them was *PPARG* which has been associated with metabolic diseases like diabetes and obesity. Because of this, the objective of this work was to characterize the selection process suffered by the variants of the *PPARG* gene in an indigenous population and to determine if these alleles are associated with specific transcriptional profiles in adipose tissue. Our outcomes show these variants have formed a specific haplotype for indigenous populations of Mexico with 35 variants distributed mainly in intron regions and some of them in regulatory regions, and this haplotype could have been selected 15,000 years ago. This outcome agrees with the estimated time in which the human migration to the Americas could occur. This event occurred in unfavorable environmental conditions like extremately cold climates, poor availability of food and the need to travel long routes on foot. Taking into consideration that this haplotype is in *PPARG*, which regulates several important processes related with energy metabolism like adipogenesis, our outcomes suggest this migration event could be the selective pressure that contributed to the selection of these variants. Furthermore, through differential expression analysis in adipose tissue, we identified 281 differentially expressed transcripts in carrying individuals of the haplotype under selection. Some of these transcripts are important for biological processes related to the lipid metabolism. Given we identified that some of the variants into the haplotype under selection on *PPARG* are in regulatory regions of this gene, together with results of genomic differential analysis, we infer this haplotype under selection could be changing the regulation of expression of the gene which in turn modifies the

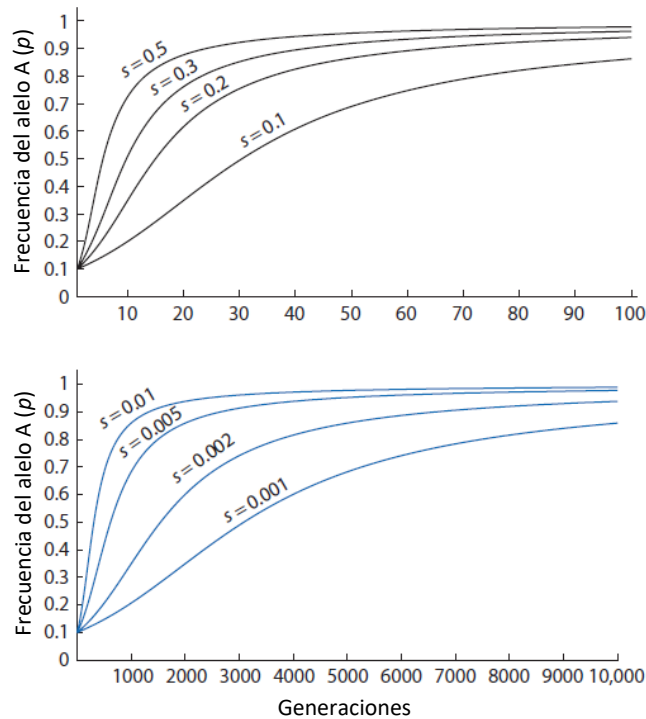
expression of the genes regulated by *PPARG*. In consequence, there is an expression signature that could result in a more effective management in the energy of the organism of the first settlers of the Americas.

## II. Introducción

### 2.1 Selección natural

Existen cuatro fuerzas que actúan en conjunto o de forma separada sobre los procesos evolutivos en las poblaciones: la deriva génica, el flujo génico, la mutación y la selección natural (Saeb & Al-Naqeb, 2016). La deriva génica es un cambio aleatorio en la frecuencia de los alelos (forma alternativa de un gen) que contribuyen a la variabilidad de la siguiente generación como consecuencia de procesos estocásticos. Mediante esta fuerza evolutiva, un alelo dentro de la población puede tener un incremento o una disminución drástica en su frecuencia de una generación a otra, e incluso ser eliminado hasta que aparezca de nuevo ya sea por mutación o flujo génico (Hamilton, 2009; Masel, 2011). El flujo génico es el movimiento de alelos de una población a otra a través de eventos de migración en una población y posterior apareamiento con los individuos de otra. El flujo génico juega un papel importante en la diversidad génica, ya que una migración limitada entre dos poblaciones puede conducir a la diferenciación génica entre estas poblaciones (Saeb & Al-Naqeb, 2016; Tigano & Friesen, 2016). La mutación se refiere a la aparición de nuevos alelos en una población debido a errores de replicación del ADN durante la mitosis o meiosis (Hamilton, 2009). Por otra parte, la selección natural es un proceso que ocurre cuando los individuos de una población difieren en su eficacia biológica o “fitness” para sobrevivir o reproducirse, como consecuencia de la aparición de alelos que aumentan dicha eficacia (Stern & Nielsen, 2019). Generalmente, estos alelos se adquieren a través del tiempo por procesos mutagénicos a lo largo del genoma de una especie. Con el tiempo los alelos que son favorables, es decir, los que tuvieron mayores probabilidades de ser heredados a las siguientes generaciones, se fijan en la población incrementando su frecuencia alélica y generando una adaptación de la población a ciertos ambientes (adaptación local) (Hamilton, 2009; Tiffin & Ross-Ibarra, 2014). En algunos casos, los fenotipos con adaptación local interactúan de formas diferentes con otros ambientes, siendo más benéficos o incluso perjudiciales (Rees et al., 2020). Para conocer la intensidad con la que está operando esta fuerza evolutiva sobre la

frecuencia de los alelos seleccionados a lo largo de las generaciones se utiliza el coeficiente de selección ( $s$ ), el cual se calcula utilizando la diferencia entre la media de la eficacia biológica de un genotipo de referencia y la del genotipo en estudio (**Fig. 1**) (Thurman & Barrett, 2016).

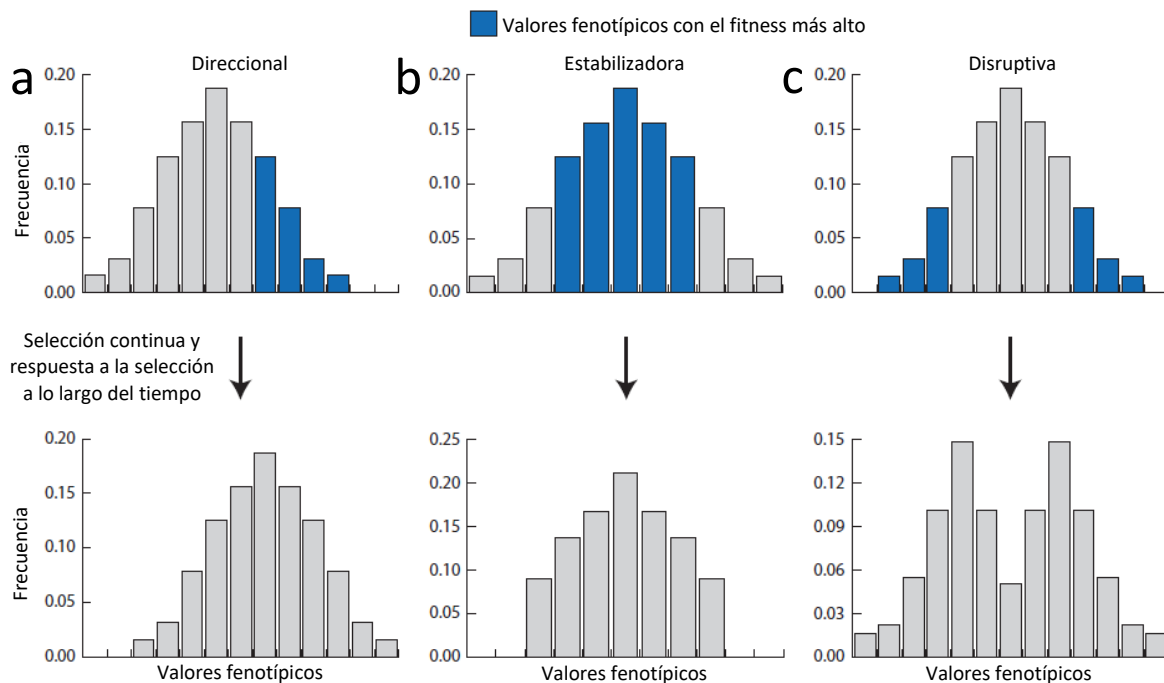


**Figura 1.- La fuerza de la selección natural influye en la tasa de cambio en las frecuencias genotípicas y alélicas.** En esta ilustración, la selección actúa contra el homocigoto recesivo ( $aa$ ). En el panel superior se muestra una selección natural intensa en donde la viabilidad del genotipo  $aa$  es 10-50% menor que el de otros genotipos. El panel inferior muestra una selección natural débil en donde la viabilidad del genotipo es 1-0.1% menor que el de otros genotipos. Se puede notar una amplia diferencia entre la escala de tiempo de ambas gráficas, en donde a menor coeficiente de selección, mayor es el número de generaciones que tarda la variante seleccionada en incrementar su frecuencia. Tomado y traducido de *Population Genetics* (p.199), por Matthew B. Hamilton, 2007, Wiley-Blackwell.

Existen tres tipos de selección natural: selección direccional, selección estabilizadora y selección disruptiva (**Fig. 2**). La selección direccional o selección positiva, ocurre cuando uno de los extremos de la distribución genotípica de la



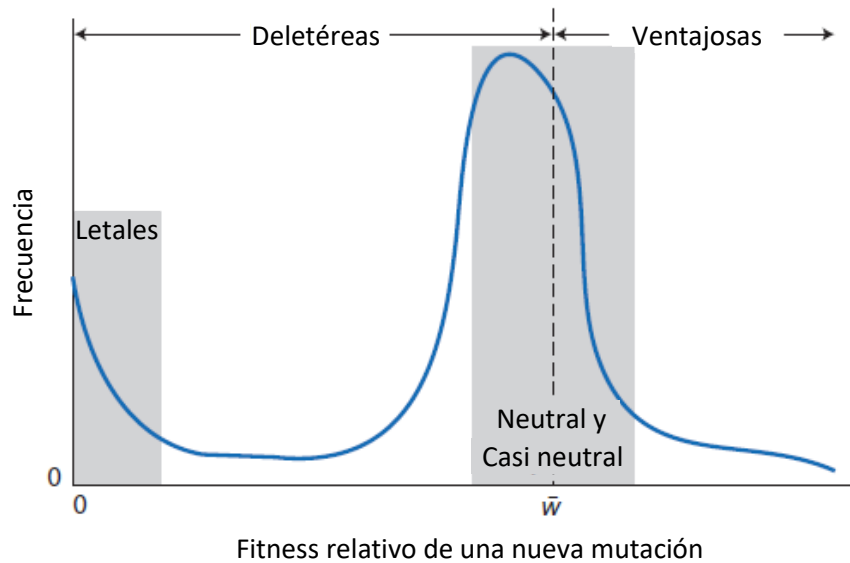
población es el que tiene la eficacia biológica más alta y por lo tanto es la que tendrá un incremento en la frecuencia del alelo a través de las generaciones (**Fig. 2a**). En la selección estabilizadora la mayor eficacia biológica se observa en los genotipos centrales de la distribución y, por el contrario, en la selección disruptiva las eficacias biológicas más altas son los genotipos de ambos extremos (**Fig. 2b y 2c**)(Hamilton, 2009).



**Figura 2.- Tipos de selección natural en rasgos cuantitativos.** El panel superior muestra la distribución inicial. El panel inferior muestra la distribución fenotípica después de haber ocurrido la selección natural. **(a)** Selección direccional ocurre cuando los valores genotípicos del extremo inicial o final de la distribución tienen el fitness más alto. **(b)** La selección estabilizadora ocurre cuando los rasgos intermedios tienen el fitness más alto. **(c)** La selección disruptiva es cuando los rasgos en los extremos de la distribución genotípica tienen el fitness más alto. La respuesta a la selección direccional aumenta o disminuye la media genotípica poblacional. La respuesta a la selección estabilizadora o disruptiva no cambia la media, pero disminuye o aumenta la varianza de la distribución genotípica. Tomado y traducido de *Population Genetics* (p.299), por Matthew B. Hamilton, 2007, Wiley-Blackwell.

## 2.2 Variación genética

La variación genética son todas aquellas diferencias en la secuencia de ADN que existen entre los individuos de una misma especie (Ellegren & Galtier, 2016). La variación genética se encuentra compuesta de variantes de un solo nucleótido (SNVs), pequeñas inserciones y deleciones, repeticiones cortas en tándem, inserciones de elementos transponibles, inversiones y variaciones en el número de copias (CNVs) (Perry, 2008). Estas variaciones en la secuencia del ADN, pueden estar localizadas en regiones codificantes o no codificantes del genoma y que, por lo tanto, tienen diferentes consecuencias sobre la eficacia biológica relativa de un individuo, pudiendo ser catalogadas como letales, deletéreas o perjudiciales, ventajosas y neutrales, presentándose en diferente frecuencia cada una de estas categorías (**Fig. 3**) (Quinn et al., 2019). Un alelo letal es aquel que conduce a la muerte, reduciendo la probabilidad de alcanzar la etapa reproductiva y por lo tanto la probabilidad de tener descendencia, uno deletéreo reduce la probabilidad de supervivencia o éxito de reproducción porque el alelo produce un fenotipo nocivo que reduce la eficacia biológica de la población frente al ambiente en el que se desarrolla, mientras que con un alelo ventajosos ocurre lo inverso a lo anterior, es decir, dentro del mismo contexto ambiental el alelo incrementa la eficacia biológica de la población y finalmente, en un alelo neutral, las mutaciones tienen pequeños o nulos efectos positivos o negativos y por lo tanto no influyen en la eficacia biológica de la población (Casillas & Barbadilla, 2017; Hamilton, 2009). La variación genética también se puede interpretar como el resultado del balance entre la aparición y desaparición de variantes genéticas en una población (Ellegren & Galtier, 2016).



**Figura 3.- Distribución hipotética de los efectos de las mutaciones en el fenotipo que finalmente impactan en el fitness del genotipo.** Las mutaciones que tienen un fitness medio menor que el fitness de la población ( $\bar{w}$ ) disminuye su frecuencia a causa de la selección natural (deletéreas). El área sombreada alrededor de  $\bar{w}$  indica la zona en donde las mutaciones tienen pequeños efectos en el fitness relativo al efecto de la deriva génica (el ancho de la zona neutral depende del tamaño efectivo de la población). El área sombreada cerca del fitness medio igual a 0 indica las mutaciones que causan fallas en la reproducción o que son letales. Las mutaciones letales son más comunes desde que son una categoría que incluye muchos grados de severidad resultantes de diversas causas. Los efectos en el fitness de las mutaciones son inherentemente difíciles de medir debido a la rareza de los eventos de mutación, el pequeño efecto de muchas mutaciones y la dependencia del fitness al contexto ambiental. Tomado y traducido de *Population Genetics* (p.158), por Matthew B. Hamilton, 2007, Wiley-Blackwell.

Por otra parte, existen tres hipótesis respecto al cómo se ha mantenido la variabilidad genética en las poblaciones a través de la selección natural conocidas como clásica, balanceadora y neutral. En la hipótesis clásica, la variación genética es poco común y está conformada principalmente por nuevos alelos mutantes con efecto dañino en el fenotipo, por lo que la selección natural las mantiene en baja frecuencia. En la hipótesis balanceadora, la variación genética es abundante y mantenida por una selección natural estabilizadora y los individuos que más se benefician son los del genotipo heterocigoto. Por último, el modelo neutral propone que la mayoría de los alelos de la variación genética tienen poco o nulo efecto en la eficacia biológica de un organismo manteniendo así la variación genética en la

población con poca o nula participación de la selección natural (Hartl & Clark, 2007; Hedrick, 2011).

### **2.3 Desequilibrio de ligamiento**

Si bien la mutación es el proceso por el cual se generan nuevas variantes, es la recombinación genética la fuerza evolutiva que permite que estas nuevas variantes se combinen o no dentro de bloques de desequilibrio de ligamiento (LD, por sus siglas en inglés) que podrían ser posteriormente blancos de la selección natural (Auton & McVean, 2012). Estos bloques de LD, se generan durante la meiosis, donde la recombinación genética de secuencias de ADN entre dos donadores produce una nueva secuencia de ADN. (Keith, 2008). Dentro de estas secuencias se da la asociación no aleatoria de los alelos de dos o más *loci* que da como resultado un aumento en la frecuencia de la cosegregación de estas secuencias (Slatkin, 2008). Por lo tanto, cada par de alelos asociados de manera no aleatoria tiene un valor de LD diferente que es denotado con la letra D y se calcula con las frecuencias de ambos alelos. Los valores máximos y mínimos que puede adquirir D, varían en función de la frecuencia de los alelos, lo que dificulta que este valor sea comparado en un par de alelos con el de otros, por lo que, se han propuesto otros valores que eliminan este problema como el valor D' que es la división de D/ D máxima o el valor  $r^2$  que es el cuadrado de la correlación convencional de la frecuencia en la muestra (Sved & Hill, 2018).

En el genoma se pueden localizar segmentos con variantes genéticas en LD, estos segmentos son conocidos como bloques de LD y, al menos en el genoma humano, varían en tamaño desde algunas cuantas kilobases (kb) hasta más de 100 kb (Slatkin, 2008). Los bloques de LD pueden ser influenciados por algunas fuerzas evolutivas como la deriva génica, el flujo génico, la selección natural, la endogamia o la tasa de recombinación (Hedrick, 2011). En el caso de la selección natural, si una variante dentro del bloque de LD se encuentra bajo selección positiva y por lo tanto aumentando de frecuencia, podría conducir al efecto conocido como autostop genético a lo largo del bloque de LD. El autostop genético es el aumento de la

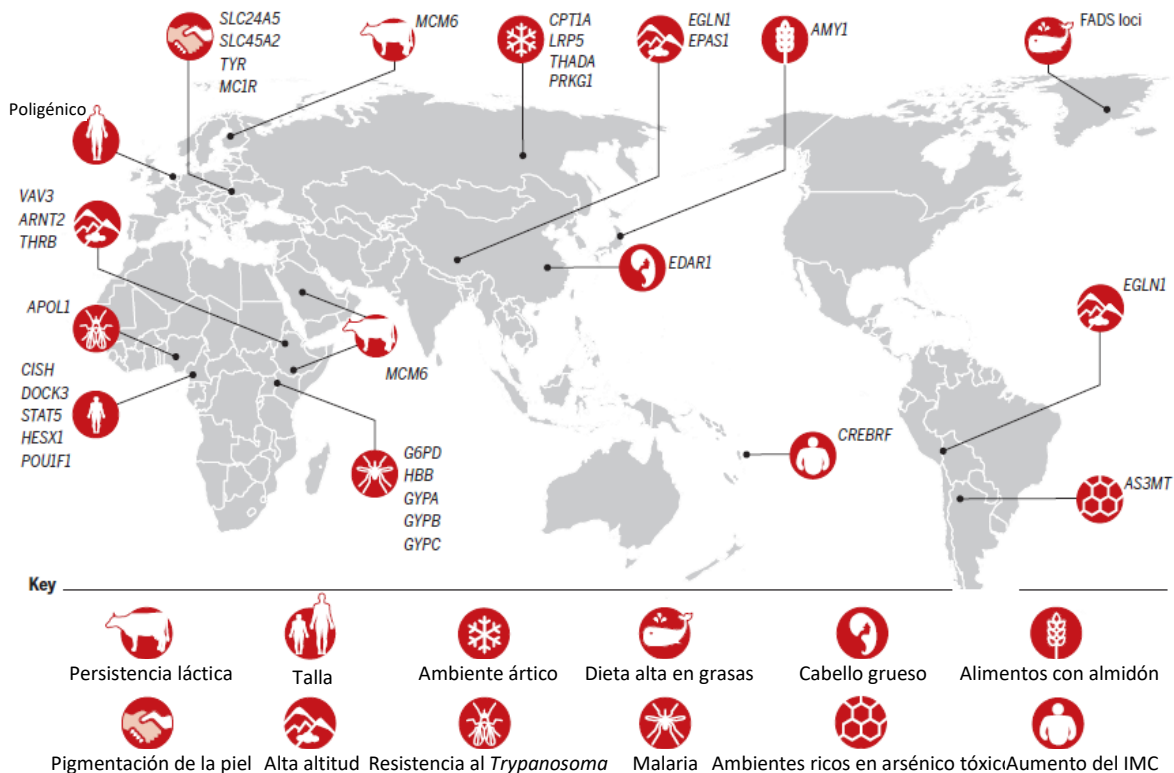
frecuencia del resto de las variantes dentro de un bloque de LD, aunque tengan un impacto neutro en el fenotipo, en consecuencia al aumento en la frecuencia de la variante bajo selección dentro del mismo bloque de LD (Slatkin, 2008). El estudio del autostop genético ha llevado al desarrollo de métodos para detectar los sitios en el genoma humano con probable evidencia de algún procesos de selección natural, por ejemplo el estadístico de homocigosidad haplotípica extendida que mide el decaimiento del LD para identificar sitios bajo selección positiva reciente (Sved & Hill, 2018).

Por lo tanto, el LD es de gran importancia en el estudio de la genética humana porque puede reflejar si nuestro genoma ha sido influenciado por fuerzas evolutivas. Es por ello que se espera que, con los avances y disminución de precios en las tecnologías de genotipificación y un aumento de personas que decidan participar en los estudios genéticos, se logre obtener una mayor comprensión de los patrones de LD y por lo tanto de la historia de diferentes regiones del genoma humano y de sus marcadores ligados a otros alelos para aumentar el potencial de su aplicación en el entendimiento de las bases genéticas de las enfermedades complejas (Slatkin, 2008; Sved & Hill, 2018).

## **2.4 Selección natural en poblaciones humanas**

En los últimos años se han desarrollado métodos, como los estadísticos PBS (*Population Branch Statistics*) e iHS (*integrated Haplotype Score*) para buscar variantes con evidencia de selección positiva asociadas a ciertos fenotipos, tales como dieta alta en grasa o a periodos largos de ayuno (rs4988235), la pigmentación de la piel (rs1426654), la respuesta inmune (rs9272426) y el almacenamiento de energía (rs1438307) (**Fig. 4**) (Kern & Hahn, 2018; Lin et al., 2018; Lindo et al., 2016; Mathieson & Mathieson, 2018; Sabeti et al., 2006; L. Wang et al., 2020). Particularmente, en México, este tipo de estudios han identificado que SNVs en el gen *MUC19* y la región *MHC* han atravesado por procesos de selección en población mexicana (Reynolds et al., 2019; Zhou et al., 2016). Sin embargo, hasta donde es de nuestro conocimiento, aún son escasos los estudios de este tipo en

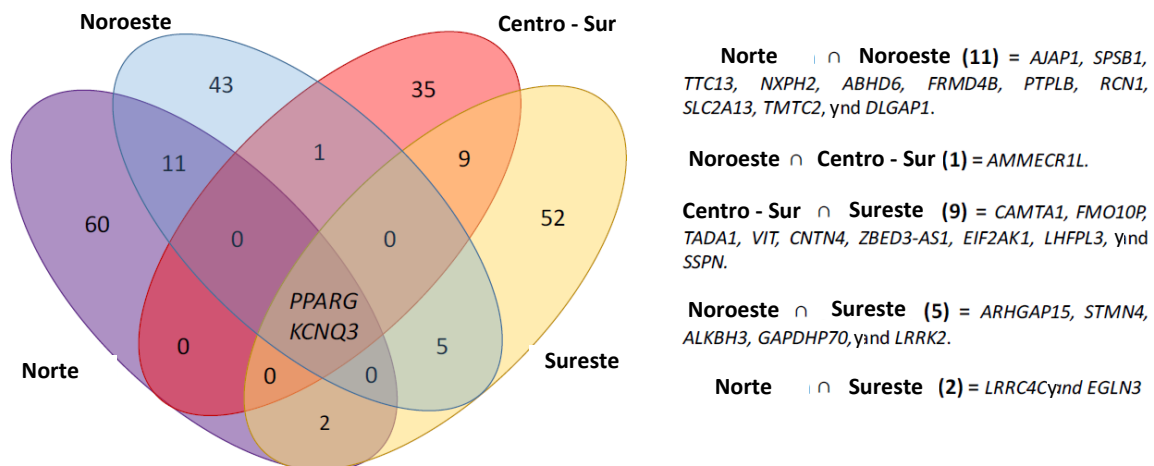
nuestra población por lo que aún se requiere de mayores esfuerzos que lleven a la detección de más variantes con señales de selección con la finalidad de entender la arquitectura genética humana.



**Figura 4.- Ejemplos de adaptaciones locales humanas.** Cada punto marca la población humana en la que se ha llevado a cabo estudios en la búsqueda de evidencia de procesos de selección. Cada ejemplo está etiquetado por el fenotipo y/o presión selectiva y los *loci* genéticos bajo selección. Tomado y traducido de “Going global by adapting local: A review of recent human adaptation” (p. 54), por S, Fan, et al., 2016, Science, 354(6308).

Recientemente, un estudio realizado previamente por nuestro grupo de trabajo, que analizó una población de 3294 indígenas (cohorte MAIS: del inglés “Metabolic Analysis in an Indigenous Sample”) provenientes de 60 de los 68 grupos étnicos de todo el país, identificó que se podían agrupar a estos individuos en 4 regiones geográficas a lo largo de México (Sureste, Centro-Sur, Norte y Noroeste) con características genéticas diferentes identificadas a través de análisis de componentes principales. Como uno de los principales resultados, en cada una de

estas regiones, se identificaron genes que contenían variantes que se encuentran bajo procesos de selección natural en el genoma de estos individuos, además de los genes con variantes con evidencia de selección en común entre estas regiones (**Fig. 5**). De manera interesante, solo dos genes (*PPARG* y *KCNQ3*) se identificaron con variantes con señales de selección en las 4 regiones geográficas que se identificaron en la población indígena mexicana (**Fig. 5**) (García-Ortiz et al., 2021). De estos dos genes, *PPARG* resalta por tener un papel clave en la diferenciación de adipocitos y en el mantenimiento de la homeostasis de la glucosa y lípidos, además de que variantes genéticas en este gen se han asociado con la susceptibilidad a enfermedades metabólicas como la diabetes y la obesidad (Janani & Ranjitha Kumari, 2015). Derivado de este estudio, se identificaron cuatro SNVs en *PPARG* con señales significativas de selección en alto LD ( $r^2 \geq 0.50$ ; **Tabla 1**) (García-Ortiz et al., 2021).



**Figura 5.- Señales compartidas de selección a través de poblaciones indígenas de diferentes regiones de México.** Diagrama de Venn que muestra el número total de genes con al menos un SNV dentro del top 1% del cruce de valores de referencia PBS (del inglés Population Branch Statistic) e iHS (del inglés integrated Haplotype Score) en cada región y el número total de señales compartidas de selección entre cada región. Tomado y traducido de “The genomic landscape of Mexican Indigenous populations brings insights into the peopling of the Americas”, por H, García-Ortiz, et al., 2021, Nature communications.

**Tabla 1 Variantes con evidencia de selección en el gen PPARG.**

SNVs	Posición en el gen	Alelo bajo selección	Frecuencia alélica				iHS†	PBS†
			Indígenas	AFR*	EUR*	ASS*		
rs6782475(T/G)	Intrón	G	0.456	0.362	0.166	0.003	2.41	0.21
rs7626560(C/T)	Intrón	T	0.451	0.164	0.150	0.002	2.46	0.22
rs9833097(G/A)**	Región regulatoria	A	0.473	0.096	0.097	0.001	2.65	0.24
rs13090265(A/G)	Río abajo	G	0.478	0.082	0.116	0.001	2.38	0.27

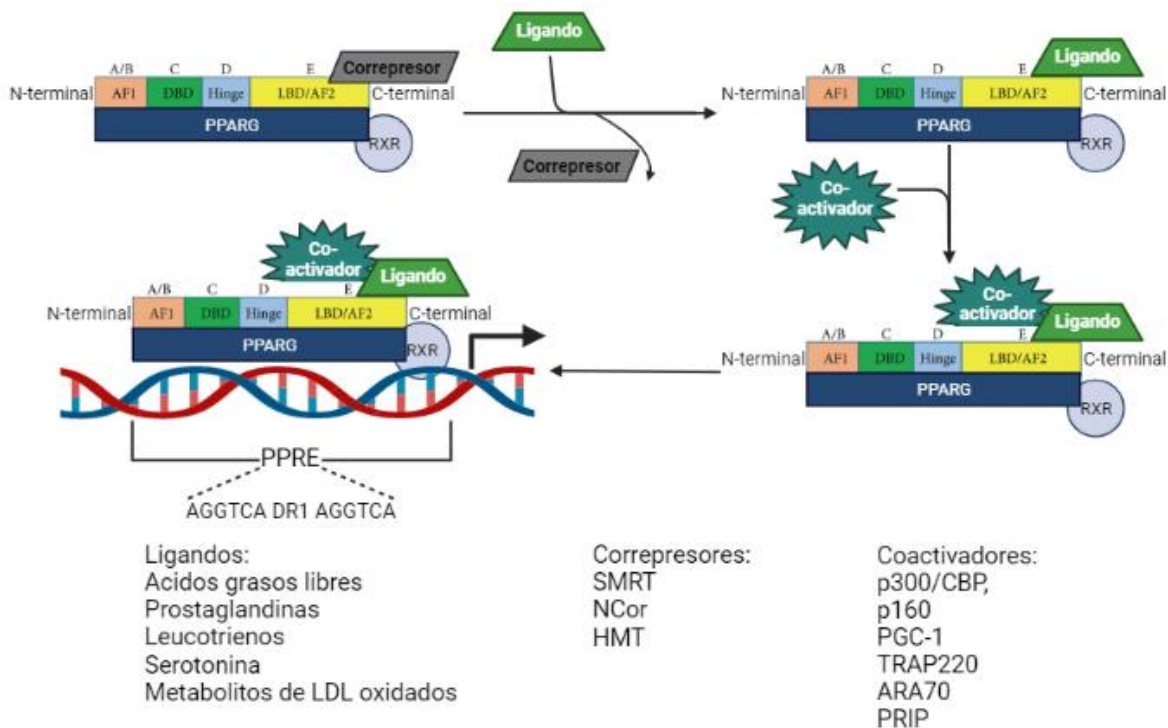
AFR: africanos; EUR: europeos; ASS: asiáticos del sur \*Basada en la frecuencia global observada en el proyecto de los 1000 genomas; †estadísticos utilizados para determinar la evidencia de selección; PBS: Population Branch statistic; iHS: integrated Haplotype Score; \*\*Variante con el valor más alto de selección; El alelo más frecuente en las poblaciones de los 1000 genomas se resalta en negritas. Nota. Adaptado de “*The genomic landscape of Mexican Indigenous populations brings insights into the peopling of the Americas*” por H, García-Ortiz, et al., 2021, Nature communications.

## 2.5 Gen PPARG

El Receptor Gamma Activado por Proliferadores Peroxisomales (PPARG, por sus siglas en inglés) forma parte de la familia de los PPARs junto a otros dos subtipos de receptores (PPAR alfa y PPAR beta/delta) (Janani & Ranjitha Kumari, 2015). Los PPAR son los factores de transcripción más estudiados y son codificados por tres genes diferentes (*PPARA*, *PPARB/D* Y *PPARG*) localizados en tres cromosomas distintos. Sin embargo, a pesar de formar parte de la misma familia y de tener características estructurales similares, cada una de estos 3 subtipos tiene funciones fisiológicas, expresión en tejidos y especificidad a ligandos diferentes (Costa et al., 2010). En el caso del gen *PPARG*, este se expresa principalmente en el tejido adiposo en donde es clave para procesos como la adipogénesis, balance de energía, biosíntesis y metabolismo de lípidos, así como la sensibilidad a la insulina y se ha descrito que la transcripción aberrante de este gen puede contribuir a la aparición de padecimientos como la obesidad y diabetes tipo II (Costa et al., 2010; Joffe & Houghton, 2016; Majithia et al., 2014). La estructura de la proteína PPARG consta de cinco dominios: el dominio A/B, que contiene a la función de activación independiente de ligando 1 (AF-1, del inglés ligand-independent transactivation function 1); el dominio de unión al ADN (DBD, del inglés DNA-binding



domain); la región bisagra (Hinge); el dominio de unión al ligando (LBD, del inglés Ligand-binding domain); y el dominio F (AF-2) que es el encargado de interactuar con cofactores como los miembros de la familia p160 el cual juega un papel primordial en la regulación funcional de este factor de transcripción (Liu et al., 2020; Powell et al., 2007; Wafer et al., 2017). Para que esta proteína lleve a cabo su función, es necesario que forme un heterodímero con el receptor retinoide X (RXR), el cual puede estar unido a correpresores como RIP140, que se une al dominio AF-2, el correpresor del receptor nuclear (NCoR) y el mediador silenciador para receptores de hormonas retinoicas y tiroideas (SMRT) que reclutan complejos de histona deacetilasa para compactar el ADN y así reprimir la transcripción. Estos correpresores son desacoplados debido a la unión de ligandos al dominio AF-2, lo cual permite la unión de coactivadores transcripcionales a PPARG. Por ejemplo, el ligando endógeno 15d-PGJ12 permite la interacción del heterodímero PPARG/RXR con los coactivadores SRC-1, TIF2, p/CIP, p300 y TRAP220. Los coactivadores varían dependiendo del tipo celular, por ejemplo, en los macrófagos es PU.1 y en adipocitos es C/EBP $\beta$ . El heterodímero PPARG/RXR activado es capaz de unirse al elemento de respuesta al Proliferador de peroxisomas (PPRE), el cual consta de repeticiones AGGTCA separados por un par de bases llamados DR1 junto con el elemento de especificidad AA ACT. Lo anterior, sumado a la función de los coactivadores de poseer o reclutar proteínas con actividad histona acetil transferasa, permite la unión del complejo RNA polimerasa II y así el inicio de la transcripción de los genes objetivo de PPARG (**Fig. 6**) (Costa et al., 2010; Powell et al., 2007).



**Figura 6.- Mecanismo de acción de PPARG.** PPARG funciona como un heterodímero obligado con el receptor retinoide X. En su estado basal, el heterodímero se encuentra asociado a un correpressor con actividad histona deacetilasa que inhibe la actividad transcripcional de PPARG. La unión de un ligando al dominio AF2 de PPARG produce un cambio conformacional en la proteína, lo que produce la liberación del correpressor. La función del ligando es reclutar a una proteína que funciona como coactivador, la cual tiene actividad histona acetilasa, lo que permite que el ADN sea accesible para el complejo de RNA polimerasa II y así inicie la transcripción de los genes blanco de PPARG. RXR: Receptor retinoide X; PPRE: Elemento de respuesta al Proliferador de peroxisomas; LDL: Lipoproteína de baja densidad; SMRT: Silenciador mediador para el receptor de hormonas retinoicas y tiroideas; NCor: Correpressor del receptor nuclear; HMT: Histona metil transferasa; PGC-1: Coactivador 1 de PPARG; TRAP220: Proteína 220 asociada al receptor de hormona tiroidea; ARA70: Proteína asociada al receptor de andrógeno; PRIP: Proteína de interacción con PPARG.

El gen *PPARG* tiene mecanismos de regulación a nivel de expresión, ya sea por factores de transcripción, incluyendo al mismo *PPARG*, por microRNAs o por cambios epigenéticos. Dentro de estos cambios epigenéticos se encuentran las marcas de histonas como la acetilación de las lisinas 9 y 27 de la histona H3 (H3K9ac y H3K27ac, respectivamente) y la di- y tri-metilación de la lisina 4 de la histona H3 (H3K4me2/3). Las modificaciones en los niveles de expresión de *PPARG*

son relevantes debido a que se ha demostrado que una reducción del 25% de su expresión podría reducir la masa y grasa corporal, mejorar la sensibilidad a la insulina y reducir el riesgo a padecer dislipidemias (Broekema et al., 2019; Powell et al., 2007).

Finalmente, también se ha mostrado que SNVs en regiones regulatorias podrían modificar la expresión de *PPARG* y causar fenotipos asociados a otras enfermedades. Por ejemplo, estudios *in vitro* han identificado SNVs cercanos a la región promotora de *PPARG* asociados con aumento del peso corporal y de la sensibilidad a la insulina. Sin embargo, debido a que no es del todo claro el mecanismo por el cual ocurre este fenómeno, son necesarios más estudios enfocados a este tipo de variantes (Costa et al., 2010; Muller et al., 2003). También se han reportado variantes que se encuentran en otras regiones importantes como la variante Pro115Gln en el dominio AF-1 que produce que *PPARG* se encuentre activa de forma constitutiva, esta variante se ha asociado a un incremento en el IMC de individuos obesos. Por otra parte, la variante Ser289Cys que podría estar provocando la formación de un puente entre Cys289 y Cys285 en el dominio LBD en *PPARG*, lo que evita la unión de correpresores a este factor de transcripción, se asocia al riesgo a desarrollar dislipidemias, hipertensión y sobrepeso (Capaccio et al., 2010; Heikkinen et al., 2007; Ristow et al., 1998).

### III. Justificación

La selección natural es un proceso que ha participado en la historia evolutiva de distintas especies, incluyendo al *Homo sapiens* desde su salida del continente africano y su dispersión por el mundo hace aproximadamente 100,000 años (Fan et al., 2016). En los últimos años se han llevado a cabo esfuerzos por desarrollar herramientas que permitan identificar huellas de esta fuerza evolutiva en el genoma humano. Como resultado, se han identificado distintos genes que podrían estar asociados a características ventajosas para poblaciones humanas específicas, incluyendo características relacionadas al metabolismo energético (Mathieson & Mathieson, 2018; L. Wang et al., 2020). Recientemente, un estudio realizado en nuestro grupo evidenció que en la población indígena mexicana existen SNVs que pudieron haber estado bajo proceso de selección natural, dentro de las cuales se encuentran variantes localizadas en *PPARG* (García-Ortiz et al., 2021). *PPARG* se expresa principalmente en tejido adiposo, donde juega un papel importante en procesos metabólicos como la adipogénesis, biosíntesis y metabolismo de lípidos y sensibilidad a la insulina a través de la regulación de la expresión de genes adipogénicos (Costa et al., 2010; Joffe & Houghton, 2016). Este gen es de gran relevancia médica debido a que ha sido asociado a padecimientos metabólicos como la diabetes y la obesidad, los cuales, para el 2005, ya formaban parte de los principales desafíos para el sistema mexicano de salud, observándose, desde entonces, un continuo incremento en la prevalencia de estos padecimientos hasta hoy en día (Janani & Ranjitha Kumari, 2015; Secretaria de Salud, 2005; Shamah-Levy et al., 2020).

Ambas enfermedades son consideradas como enfermedades multifactoriales o complejas debido a que no tienen una única causa genética o ambiental, lo que complica su estudio (Artasensi et al., 2020; Rohde et al., 2019). Por lo tanto, el estudio de estas enfermedades requiere de un enfoque multidisciplinario (Duarte et al., 2014). En este sentido, la medicina evolutiva proporciona un enfoque en el que se analiza como las variantes genéticas han interactuado con el ambiente, las fuerzas evolutivas y otros genes a lo largo del tiempo para influir sobre el genotipo

que subyace a la enfermedad y así aportar al entendimiento causal de las enfermedades actuales (Romero & Farias, 2014).

De esta manera, se espera que, al combinar el entendimiento actual de la dinámica evolutiva, la diversidad genética regional de México y las posibles consecuencias que los alelos bajo selección causen sobre fenotipos particulares, logremos profundizar en el conocimiento de las enfermedades que afectan a la población mexicana y se genere el potencial para contribuir a originar mejores estrategias epidemiológicas y de intervención que ayuden a mejorar la salud de nuestra población.

## **IV. Objetivos**

### **4.1 Objetivo general**

Estimar el proceso de selección que sufrieron las variantes del gen *PPARG* en población indígena y determinar si estos alelos se asocian con perfiles transcripcionales específicos en tejido adiposo.

### **4.2 Objetivos Específicos**

- Identificar la estructura del haplotipo del gen *PPARG* con alelos con evidencia de selección (rs6782475\*T, rs7626560\*T, rs9833097\*A, rs13090265\*A) en una muestra de 76 genomas indígenas genotipificados a nivel de genoma completo.
- Establecer el coeficiente de selección, así como estimar el tiempo en el cual el haplotipo derivado empezó a incrementar su frecuencia utilizando estrategias bioinformáticas.
- Evaluar los perfiles transcripcionales en tejido adiposo entre portadores del haplotipo derivado (con evidencia de selección) y portadores del haplotipo de referencia.

## V. Metodología

### 5.1 Población de estudio

Las muestras incluidas en este estudio pertenecen al biobanco del Laboratorio de Inmunogenómica y Enfermedades Metabólicas del Instituto Nacional de Medicina Genómica (LIEM-INMEGEN).

Para identificar la estructura del haplotipo bajo selección de *PPARG* y estimar el tiempo en el cual se empezó a incrementar su frecuencia en la población indígena mexicana, se utilizó una submuestra de la cohorte MAIS “Metabolic Analysis in an Indigenous Sample”. Esta submuestra estuvo conformada de 76 indígenas mexicanos pertenecientes a 27 grupos étnicos del país, con secuenciación del genoma completo por medio de la plataforma Illumina. De todos ellos, también se cuenta con la información clínica, datos antropométricos (talla, peso y composición corporal), bioquímicos (glucosa, colesterol total, colesterol HDL y triglicéridos) y presión sanguínea (Mendoza-Caamal et al., 2020).

Para la evaluación de los niveles de transcripción, se utilizó una población integrada por 12 individuos mestizos mexicanos con un índice de masa corporal (IMC)  $\geq 35$  kg/m<sup>2</sup> que fueron sometidos a cirugía bariátrica en la Clínica integral de cirugía para la obesidad y enfermedades metabólicas en el Hospital General de Tlahuac, Ruben Leñero y Ángeles del Pedregal de la Ciudad de México (Barajas-Olmos et al., 2018). De ellos se recolectó una muestra de tejido adiposo visceral durante una operación bariátrica a la que fueron sometidos. De todos ellos se cuenta también con su información genética (Illumina, HumanHT-12 v4 Expression BeadChips), bioquímica (glucosa, hemoglobina glicosilada, insulina, colesterol total, colesterol HDL, colesterol LDL y triglicéridos), además de medidas antropométricas y de presión arterial (Barajas-Olmos et al., 2018)

El presente trabajo se desarrolló acorde a la declaración de Helsinki y fue aprobado por el comité de ética en investigación del INMEGEN (número de protocolo: CEI 2019/08).

## 5.2 Identificación de la estructura de los haplotipos bajo selección

Se utilizaron los archivos en formato Plink (.bed, .bim y .fam) provenientes de los 76 genomas completos de indígenas para identificar la estructura del haplotipo de las SNVs bajo selección en el gen *PPARG* (haplotipo bajo selección) (**Tabla 1**) utilizando el software Plink v1.9. Se consideraron en desequilibrio de ligamiento aquellas variantes con un valor de  $r^2 \geq 0.50$  con respecto a la variante con los valores más altos de selección rs9833097 (**Tabla 1**). La búsqueda de variantes en alto LD se realizó en una región de 158 kb que abarca desde 1 kb antes del inicio de la transcripción del gen *PPARG* hasta 1 kb posterior a la ubicación de la última variante del bloque de LD, que fue la variante rs13090265. La visualización del haplotipo bajo selección se realizó con el software Haploview v4.2. Posteriormente se comparó la estructura del haplotipo bajo selección en la población indígena con el de las poblaciones de la fase tres del proyecto 1000 genomas (The 1000 Genomes Project Consortium, 2015), siguiendo la misma estrategia de análisis descrita anteriormente.

## 5.3 Anotación funcional *in silico* de las variantes que componen el haplotipo bajo selección en *PPARG*

Para evaluar el posible impacto funcional de las SNVs que componen el haplotipo bajo selección se construyó un archivo en formato vcf (variant calling format) a partir de los archivos en formato ped (PLINK/Haploview text pedigree + genotype table) utilizando el software Plink v1.9. El archivo vcf se ingresó a la plataforma web *Variant Effect Predictor* (VEP) (Yates et al., 2019) para determinar, *in silico*, las posibles consecuencias a nivel funcional de las variantes que componen el haplotipo.

También se evaluó si las variantes del haplotipo bajo selección se encuentran dentro de regiones importantes para la regulación de *PPARG*. Para esto, se analizaron las interacciones del promotor de este gen con otras regiones regulatorias alrededor de



él y los resultados se compararon con la ubicación de las variantes del haplotipo bajo selección, utilizando la base de datos web Gene Expression Omnibus (<https://www.ncbi.nlm.nih.gov/geo/>) para extraer los datos de captura del promotor por HiC (del inglés High-throughput chromosome conformation capture) de pre adipocitos blancos diferenciados a adipocitos *in vitro* (GSM3004355). De los datos procesados en esta base de datos, se obtuvieron las coordenadas de interacción y se convirtieron a formato INTERACT para su visualización en una gráfica de interacciones arco utilizando el UCSC genome browser ([http://genome.ucsc.edu/cgi-bin/hgGateway?hgsid=1139515969\\_K2j87Q5nP7apBNa8yW0KAh39TpVV](http://genome.ucsc.edu/cgi-bin/hgGateway?hgsid=1139515969_K2j87Q5nP7apBNa8yW0KAh39TpVV)). Se analizó una ventana de 500 kb en la cual las interacciones se compararon con los picos de accesibilidad de la cromatina medidos por ATAC-seq, DNAsa I, marcas de cromatina (H3K27ac, H3K4me1 y H3K4me3) y los SNVs que integraron el haplotipo bajo selección. Para los picos de accesibilidad se utilizaron datos procesados de tejido adiposo humano (GSM3331037, <http://cistrome.org/db/#/>). Los datos de DNAsa I corresponden a 125 tipos celulares mientras que las marcas de cromatina a 7 líneas celulares, ambos grupos de datos del proyecto ENCODE. Finalmente, se mapeo la posición de los probables elementos regulatorios con los que puede interactuar *PPARG* (GeneHancer, ENCODE) para comparar los resultados de las interacciones arco y la posición de las variantes del haplotipo bajo selección.

#### **5.4 Determinación del coeficiente de selección del haplotipo de *PPARG***

Para la determinación del coeficiente de selección ( $s$ ) se utilizó el modelo propuesto por Zhou et al. (2016) (**Ec. 1**). Este modelo utiliza la frecuencia alélica de la variante bajo selección en la población de interés (que en este estudio es la población indígena;  $P_0$ ), la frecuencia de la misma variante en una población de referencia ( $P_1$ ) y el tiempo aproximado en generaciones en que se estima que estas dos poblaciones divergieron. Como poblaciones de referencia se utilizaron los datos genómicos de tres poblaciones diferentes: la población Tagar ( $n=8$ ) y la población Koriak ( $n=11$ ), ambas siberianas y fueron seleccionadas por encontrarse dentro de

la posible ruta que el humano moderno tomó a su salida de África para llegar al continente de América (Damgaard et al., 2018; Lazaridis et al., 2014; Raghavan et al., 2015). La tercera población fue la población peruana de Lima (PEL) del proyecto 1000 Genomas (n=85) (The 1000 Genomes Project Consortium, 2015), que es una población posterior al asentamiento de la población indígena mexicana. Para el cálculo del tiempo de divergencia en generaciones de estas poblaciones con la indígena mexicana se dividió el tiempo aproximado en años en que divergieron entre el tiempo que se estima que dura una generación humana (29 años) (Ewbank, 2016). El tiempo de divergencia entre las poblaciones siberianas y la americana que se utilizó fue de 23 mil años y para la población peruana de 2,100 años (Raghavan et al., 2015; Harris et al., 2018).

$$s = \exp(\log(p_1/p_0)/G) - 1$$

**Ecuación 1.- Modelo simple para estimar el coeficiente de selección.** s: Coeficiente de selección; p1: Población de referencia; p0: Población de interés; G: Tiempo de divergencia de las dos poblaciones en generaciones. Tomado y adaptado de “Strong Selection at MHC in Mexicans since Admixture” (p. 13), por Q. Zhou, et al., 2016, PLOS Genetics, 12(2).

## 5.5 Estimación del tiempo de selección del haplotipo de *PPARG*

Para calcular el tiempo en que la variante de interés comenzó a incrementar su frecuencia en la población indígena, se utilizó el software starTMRCA v0.6-3 (Smith et al., 2018). Este software asume que las poblaciones comparten un mismo ancestro en común (genealogía de estrella) para determinar el tiempo al ancestro común más reciente (TMRCA) en generaciones de una secuencia mediante modelos de coalescencia. Para ello, el algoritmo de starTMRCA requiere de poblaciones de referencia para estimar el TMRCA, por lo que se incluyeron 20 individuos homocigotos para el haplotipo de referencia, seleccionados aleatoriamente de la población Yoruba (YRI), Británica (GBR), Finlandesa (FIN) y Han de China (CHB) del Proyecto 1000 genomas (5 de cada población, n = 20) (The 1000 Genomes Project Consortium, 2015). Para cada una de las simulaciones, cada corrida fue estandarizada tomando en cuenta los siguientes parámetros: tasa de

mutación ( $u$ ) =  $1.2 \times 10^{-08}$ , tasa de recombinación ( $R$ ) =  $1 \times 10^{-08}$ , iteraciones ( $i$ ) = 10,000, de acuerdo a lo propuesto previamente por S. Mathieson & Mathieson (Mathieson & Mathieson, 2018). Debido a que starTMRCA entrena al modelo que genera utilizando la información de la iteración anterior (Método de Monte Carlo basado en cadenas de Markov, MCMC por sus siglas en inglés) hasta lograr que los valores de TMRCA sean reproducibles entre una iteración y otra, de las 10,000 iteraciones totales se descartaron las primeras 6,000 que realiza el software, de acuerdo con las recomendaciones de los mismo autores. El resultado final del TMRCA para cada variante fue el promedio de las últimas 4,000 iteraciones (Mathieson & Mathieson, 2018). Por último, para conocer el TMRCA en años, el resultado se multiplicó por el tiempo en años de una generación humana  $T_g = 29$  (Ewbank, 2016). Para validar los resultados, el procedimiento se realizó 10 veces de manera independiente (bootstrapping) para cada una de las 4 variantes con señales de selección en el gen *PPARG* y se compararon los resultados a través de una prueba ANOVA de un factor con la prueba post hoc HSD Tukey.

## 5.6 Selección de muestras de tejido adiposo

Se realizó una búsqueda intencionada del haplotipo bajo selección identificado en población indígena en el genotipo previamente obtenido de 12 individuos mediante el microarreglo HumanHT-12 v4 BeadChip Illumina®. Se analizaron a los portadores del haplotipo silvestre (homocigotos para el alelo no seleccionado de las 4 variantes bajo selección) vs. portadores del haplotipo bajo selección (heterocigotos para el alelo bajo selección de las 4 variantes bajo selección). Para ello, se identificaron a los homocigotos para el haplotipo silvestre (HS;  $n = 3$ ) y heterocigotos (HE;  $n = 9$ ) del haplotipo bajo selección y se compararon sus parámetros antropométricos, bioquímicos y de presión arterial para descartar posibles variables de confusión en el análisis de expresión diferencial (**Tabla 2**).

La comparación de los niveles de expresión del tejido adiposo fue realizado con el paquete Limma (Linear Models for Microarray and RNA-seq Data) de R. Para

ello, se creó una matriz de contraste para realizar las comparaciones por parejas entre los dos grupos utilizando un modelo linear ajustado y posteriormente obtener una lista de transcritos diferencialmente expresados, tomando como criterio una tasa de cambio (FC, por sus siglas en inglés)  $> |1|$  y un valor  $p < 0.05$ . Con la lista obtenida, se realizaron análisis de redes de interacción en la plataforma de STRING (<https://string-db.org/>). Como postprocesamiento de la red, se eliminaron los transcritos que no mostraron interacción con otros y aquellos que tenían un puntaje de interacción menor a 0.900. También se realizaron análisis de enriquecimiento de vías utilizando la plataforma de Webgestalt (<http://www.webgestalt.org/>) y 3 bases de datos funcionales diferentes (*Gene Ontology* (GO) – Procesos biológicos, GO – Función Molecular y Rutas KEGG) (Kanehisa et al., 2021; Mi et al., 2019; J. Wang et al., 2017).

**Tabla 2 Características bioquímicas y antropométricas de los 12 individuos mestizos con obesidad mórbida seleccionados para el análisis diferencial genómico.**

	Total (n = 12)	HS (n=3)	HE (n=9)	p
<b>Edad (años)<sup>b</sup></b>				
M (IC95%)	39.0 (32.0 – 46.0)	41.3 (20.2 – 62.5)	38.2 (28.9 – 47.6)	1
DE	11.1	8.5	12.2	
M <sub>e</sub> (RQ)	39.0 (36.0 – 46.75)	38.0 (35.0 – )	40.0 (36.0 – 46.5)	
<b>Glc (mg/dL)<sup>b</sup></b>				
M (IC95%)	105.2 (81.0 – 129.4)	105.0 (13.9 – 196.0)	105.2 (73.9 – 136.6)	1
DE	38.1	36.66	40.8	
M <sub>e</sub> (RQ)	92.5 (84.5 – 107.25)	97.0 (73.0 – )	88.0 (85.0 – 106.5)	
<b>CT (mg/dL)<sup>b</sup></b>				
M (IC95%)	180.1 (152.7 – 207.5)	211.3 (52.5 – 370.2)	169.7 (144.8 – 194.6)	0.354
DE	43.1	64.0	32.4	
M <sub>e</sub> (RQ)	174.5 (145.5 – 202.5)	179.0 (170.0 – )	158.0 (141.5 – 195.0)	
<b>TG (mg/dL)<sup>a</sup></b>				
M (IC95%)	153.5 (115.8 – 191.2)	168.7 (79.2 – 258.1)	147.9 (95.0 – 200.7)	0.519
DE	56.1	36.0	63.2	
M <sub>e</sub> (RQ)	132.0 (104.0 – 204.0)*	170.0 (132.0 – )	116.0 (104.0 – 215.8)	
<b>C-HDL (mg/dL)<sup>a</sup></b>				
M (IC95%)	43.7 (37.0 – 50.4)	47.8 (14.4 – 81.2)	42.4 (34.7 – 50.1)	0.569
DE	10.6	13.4	10.0	
M <sub>e</sub> (RQ)	41.4 (36.8 – 50.1)	42.9 (37.5 – )	39.9 (33.9 – 49.1)	
<b>C-LDL (mg/dL)<sup>b</sup></b>				
M (IC95%)	113.02 (91.9 – 134.1)	135.7 (3.8 – 267.7)	105.4 (87.4 – 123.5)	0.482
DE	33.2	53.1	23.4	
M <sub>e</sub> (RQ)	106.8 (86.6 – 125.0)	108.9 (101.4 – )	104.6 (83.3 – 120.8)	
<b>Ins (mg/dL)<sup>a</sup></b>				
M (IC95%)	28.8 (20.99 – 36.6)	24.8 (16.8 – 32.8)	30.3 (19.0 – 41.6)	0.309
DE	11.7	3.2	13.5	
M <sub>e</sub> (RQ)	25.4 (19.16 – 43.21)*	25.4 (21.3 – )	30.2 (16.9 – 44.1)	
<b>Hb1Ac (%)<sup>a</sup></b>				
M (IC95%)	5.8 (5.5 – 6.1)	5.5 (3.0 – 7.9)	5.8 (5.5 – 6.2)	0.240
DE	0.4	0.3	0.5	
M <sub>e</sub> (RQ)	5.8 (5.37 – 6.19)*	5.5 (5.3 – )*	5.8 (5.4 – 6.2)	
<b>TAsist (mmHg)<sup>a</sup></b>				
M (IC95%)	124.4 (111.2 – 137.6)	127.7 (77.0 – 178.4)	123.3 (106.4 – 140.2)	0.772
DE	20.8	20.4	22.0	
M <sub>e</sub> (RQ)	126.5 (110.0 – 138.0)	123.0 (110.0 – )	130.0 (110.0 – 136.0)	
<b>TAdia (mmHg)<sup>a</sup></b>				
M (IC95%)	68.3 (64.2 – 72.5)	70.0 (48.3 – 91.7)	60.0 (63.0 – 72.6)	0.713
DE	6.6	8.7	6.3	
M <sub>e</sub> (RQ)	70.0 (60.5 – 73.5)	74.0 (60.0 – )	70.0 (61.0 – 71.0)	
<b>IMC (Kg/m<sup>2</sup>)<sup>a</sup></b>				
M (IC95%)	40.7 (37.4 – 44.1)	40.8 (29.5 – 52.0)	40.7 (36.3 – 45.2)	0.987
DE	5.3	4.5	5.8	
M <sub>e</sub> (RQ)	42.2 (36.4 – 45.2)	40.3 (36.5 – )	44.1 (35.3 – 45.1)	
<b>Diabetes</b>				
n (%)	8 (66.66)	2 (66.66)	6 (66.66)	

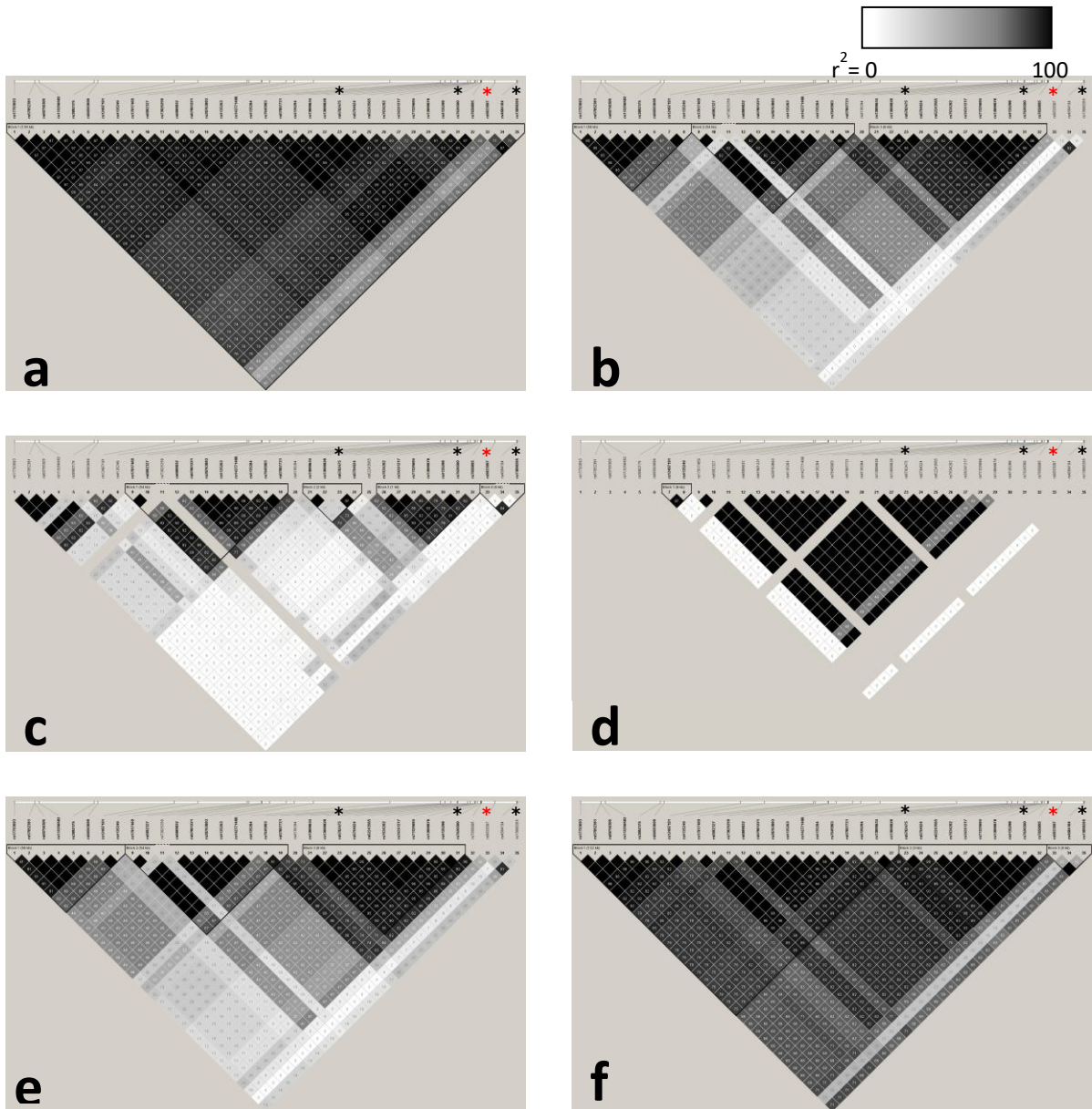
HS: Homocigotos para el haplotipo silvestre; HE: Heterocigotos para el haplotipo bajo selección; M: Media; DE: Desviación estándar; M<sub>e</sub>: Mediana; RQ: Rango intercuartílico GLC: Glucosa; CT: Colesterol Total; TG: Triglicéridos; C-HDL: Colesterol HDL; C-LDL: Colesterol LDL; Ins: Insulina; TAsist: Tensión Arterial Sistólica; TAdia: Tensión Arterial Diastólica; IMC: Índice de Masa Corporal; <sup>a</sup>: Prueba t de Student; <sup>b</sup>: Prueba U de Mann-Withney.

## VI. Resultados

### 6.1 Estructura y análisis funcional *in silico* de los haplotipos

El análisis de los patrones de LD en el gen *PPARG* utilizando a la variante rs9833097 como referencia, mostró que el haplotipo bajo selección natural está compuesto por 35 variantes con un valor de  $r^2 \geq 0.50$  entre ellas (**Fig. 7a**). De este conjunto de variantes, 32 se encontraron dentro del gen *PPARG*, desde el sitio de inicio de la transcripción hasta la secuencia de terminación (12287368 – 12434344) y 3 se encontraron río abajo: rs9833097 (12437318), rs4684104 (12443350) y rs13090265 (12443486). Al hacer la búsqueda del haplotipo identificado en la población indígena en las poblaciones del proyecto 1000 Genomas (The 1000 Genomes Project Consortium, 2015), se observó que la estructura del haplotipo en *PPARG* en estas poblaciones es diferente a la de la población indígena mexicana excepto para las poblaciones mestizas de América, sugiriendo que este haplotipo puede ser específico de poblaciones con ascendencia Nativo-americana (**Fig. 7; Anexos 1-5**).

El análisis *in silico* para la anotación funcional con la herramienta VEP de las variantes que forman el haplotipo bajo selección evidenció que el 65.7% (n = 23) de ellas tuvieron consecuencia de “Variante Intrónica”, el 28.6% (n = 10) consecuencia de “Variante en región reguladora” y el 5.7% (n = 2) consecuencia de “Variante en región Intergénica”. Dentro de las 10 variantes en región reguladora se encontró a la variante rs9833097, identificada con los valores más altos de selección (García-Ortiz et al., 2021). Las otras 9 variantes fueron: rs113256582, rs6802327, rs73025259, rs6809832, rs28763893, rs67867721, rs4135294, rs13099634 y rs13099828 (**Tabla 3**).



**Figura 7.- Estructura del Haplotipo bajo selección de PPARG en población indígena y poblaciones del proyecto 1000 Genomas.** Variantes del gen *PPARG* en LD ( $r^2 \geq 50$ ) con la variante rs9833097 identificadas en (a) población indígena mexicana ( $n = 76$ ), (b) europea (Ibéricos, España), (c) africana (Yoruba, Nigeria), (d) asiática del oriente (Han, China), (e) asiática del sur (Punjabi de Lahore, Pakistán), (f) americana (Ancestría Mexicana de Los Ángeles, Estados Unidos). El haplotipo se extiende desde la posición 12304472 hasta 12443486 (GRCh38). El \* señala a las variantes con señales de selección en el gen *PPARG* (Tabla 1); En rojo se señala la variante rs9833097.

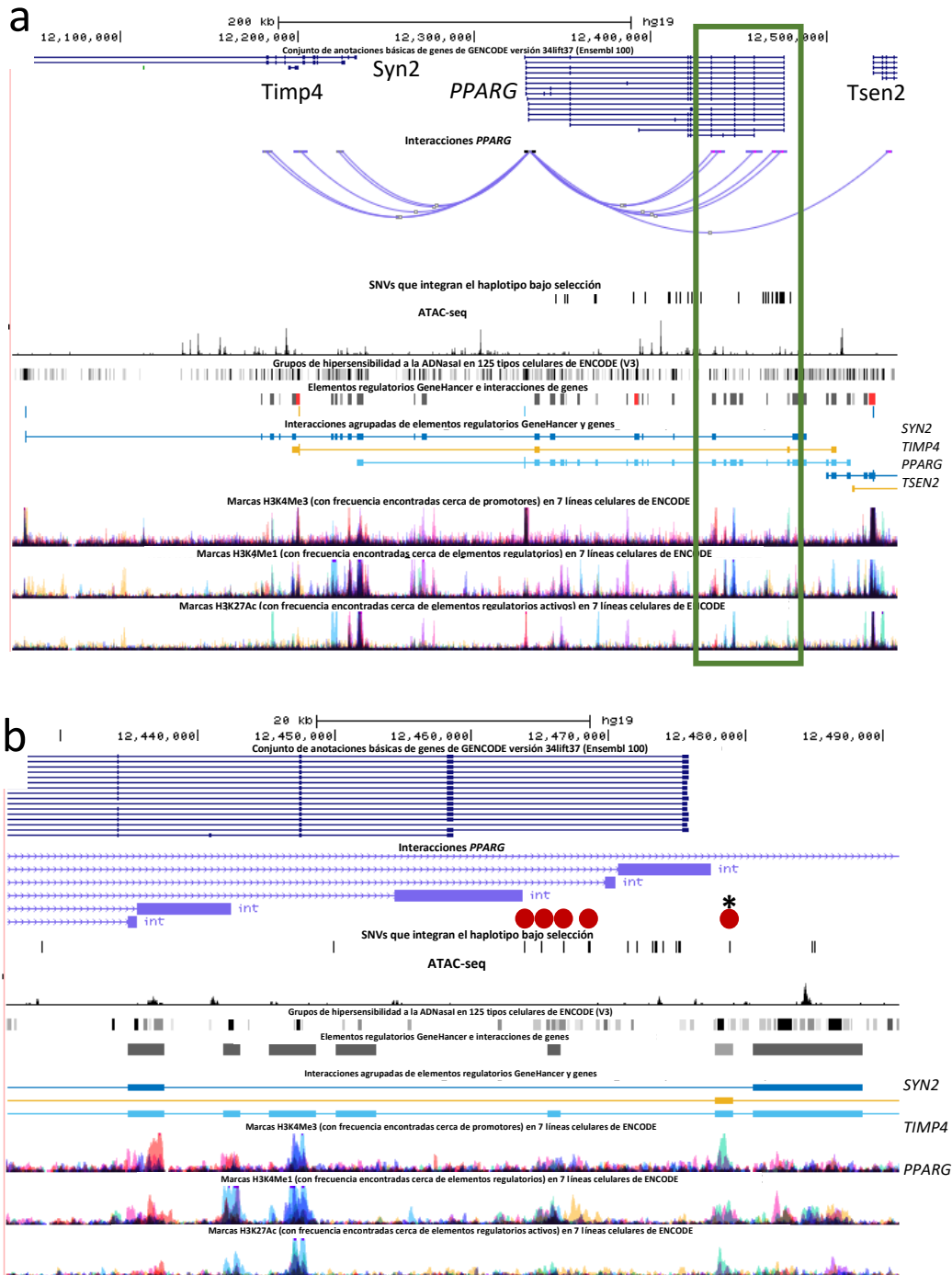
**Tabla 3.- Consecuencia de las variantes del haplotipo bajo selección (n = 35).**

Consecuencia	Variantes (%)	IDs
Variante intergénica	5.7%	rs4684104 y <b>rs13090265</b>
Variante intrónica	65.7%	rs17793693, rs57052381, rs60756305, rs3892175, rs66593699, rs12497191, rs4135249, rs17817469, rs67901221, rs4135263, rs142771488, rs4135284, rs7645903, <b>rs6782475</b> , rs6794024, rs62243565, rs7634282, rs34241217, rs71325694, rs13099078, rs4135298, <b>rs7626560</b> y rs7650895
Variante en región reguladora	28.6%	rs113256582, rs6802327, rs73025259, rs6809832, rs28763893, rs67867721, rs4135294, rs13099634, rs13099828 y <b>rs9833097</b>

Las variantes con señales de selección están resaltadas en negritas

Este resultado sugiere que las variantes que integran el haplotipo bajo selección en el gen *PPARG*, a pesar de no estar dentro de regiones codificantes del gen, de alguna manera tienen un impacto en su funcionalidad. Por lo tanto, realizamos un análisis *in silico* de interacción del promotor con regiones “enhancer” para determinar si estas variantes se encuentran en regiones reguladoras del gen *PPARG*. El análisis epigénético utilizando datos de HiC de tejido adiposo y RNA-seq de 7 líneas celulares del proyecto ENCODE (The ENCODE Project Consortium et al., 2020) mostró, que, efectivamente, en regiones con marcas de enhancer dentro del gen *PPARG* (H3K4me3 y H3K27ac) se encuentran las variantes rs67867721, rs4135294, rs13099634 y rs13099828 que fueron anotadas en regiones reguladoras por la plataforma VEP (**Fig. 8**). Además, la variante rs9833097, se encontró dentro de una región con marcas de región promotora (H3K4me1) (**Fig. 8**). Lo anterior sugiere que las variantes del haplotipo bajo selección que se encuentran en regiones reguladoras de *PPARG* que, además interactúan con el promotor, podrían causar modificaciones en la regulación de la expresión de este gen.





**Figura 8.- Interacciones Promotor-Enhancer del gen *PPARG* en datos de HiC de tejido adiposo y RNA-seq de 7 líneas celulares del proyecto ENCODE.** En la parte superior se observa la posición del gen *PPARG*. Las líneas de interacciones en arco muestran los sitios con los que interactúa el promotor de *PPARG*. Las líneas negras verticales representan la posición de las variantes del haplotipo bajo selección. Los picos en los datos de ATAC-seq representan sitios de accesibilidad en la cromatina. Los cuadros grises muestran las áreas hipersensibles a la DNasa. La obscuridad de cada cuadro es proporcional a la señal más fuerte detectada en un tipo celular. Los elementos regulatorios GeneHancer se encuentra

señalados con cuadros. Los cuadros rojos representan regiones promotoras y los grises regiones enhancers. La intensidad del color de cada cuadro indica su puntaje de confianza. Los genes con los que interactúan estas regiones se encuentran debajo con un color diferente para cada gen. Los cuadros en cada gen representan los sitios con los que interactúan. Los picos en las marcas H3K4me3, H3K4me1 y H3K27ac representan sitios de lecturas altas de estas marcas. Cada color representa a una de las 7 líneas celulares que se utilizaron. Rojo: células GM12878; amarillo: células H1-hESC; verde: células HSMM; turquesa: células HUVEC; Azul: células k562; morado: células NHEK; rosa: células NHLF. (a) Ventana de 500,000 bp (12,000,000 – 12,500,000 bp). (b) Ventana de 60,000 bp (12,430,000 – 12,490,000 bp). \*: Variante con el valor más alto de selección en el gen *PPARG* (rs9833097); círculos rojos indican a las variantes identificadas en regiones regulatorias mediante el análisis funcional *in silico* con VEP.

## 6.2 Coeficiente de selección

El  $s$  para el haplotipo bajo selección en *PPARG* se calculó utilizando como *proxi* la frecuencia alélica observada de la variante rs6782475 que forma parte del haplotipo identificado y que también mostró señales de selección en el estudio previo. Se utilizaron los datos genómicos de las poblaciones siberianas Tagar ( $n = 8$ ) y Koriak ( $n = 11$ ) y la población peruana ( $n = 30$ ) como poblaciones de referencia. La estimación del  $s$  en cada una de esas poblaciones fue de:  $s = 0.0042$  para la comparación de la población Indígena mexicana vs Peruanos;  $s = 0.0032$  para la comparación de la población Indígena mexicana vs Koriak;  $s = 0.0008$  para la comparación de la población Indígena mexicana vs Tagar (**Tabla 4**). Dado que entre menor sea el valor del coeficiente de selección mayor es el tiempo que tarda en aumentar su frecuencia un alelo (**Fig. 1**), estos resultados sugieren que el proceso de selección en las variantes del haplotipo de *PPARG* ocurrió hace un largo tiempo. Para determinar de forma más precisa lo anterior, se calculó el TMRCA entre la secuencia ancestral o de referencia y la secuencia del haplotipo bajo selección.

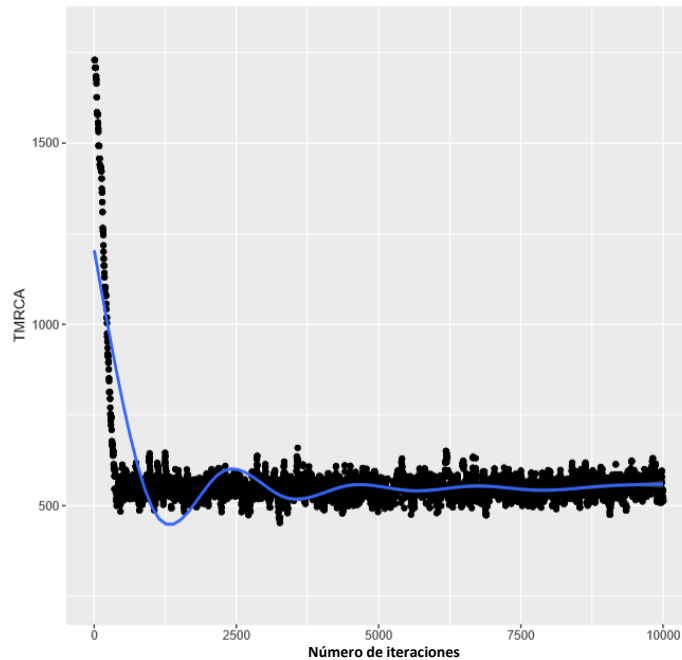
**Tabla 4.- Coeficiente de selección de la variante rs9833097 calculado a partir de la frecuencia alélica de la variante rs6782475 con la que se encuentra en desequilibrio de ligamiento.**

	FAS (rs6782475)	s
Indígenas	0.456	
Peruanos	0.324	0.0042
Koriak	0.046	0.0032
Tagar	0.25	0.0008

FAS: Frecuencia del Alelo bajo Selección;  
s: Coeficiente de Selección

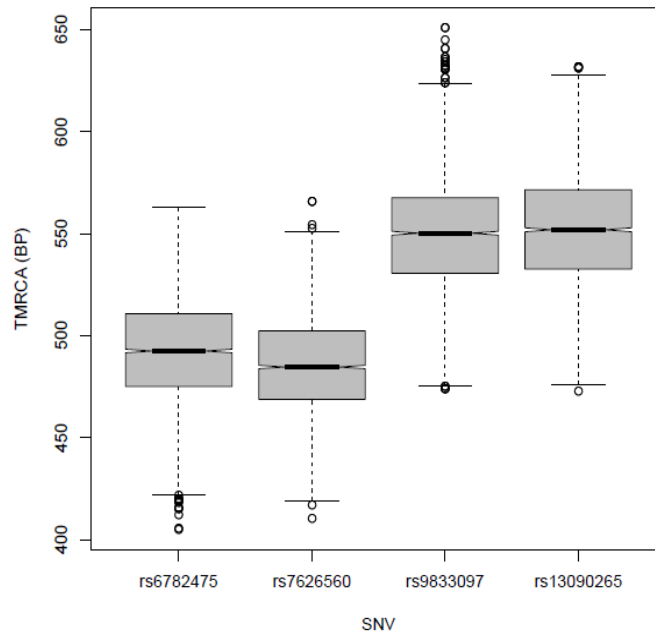
### 6.3 Tiempo al ancestro común más reciente (TMRCA)

El TMRCA se calculó de forma independiente para las cuatro variantes con señales de selección en el gen *PPARG* reportadas por García-Ortiz, et al. (2021) (**Tabla 1**) por medio de modelos de coalescencia implementados en el software StartMRCA. Para cada variante se realizaron 10 análisis independientes (bootstrapping, ver métodos), en las que se calcularon 10,000 iteraciones en cada simulación (Mathieson & Mathieson, 2018). Como se puede observar en la **Figura 9** que corresponde a los resultados obtenidos para la variante rs9833097, los resultados de las primeras iteraciones no son reproducibles, por lo que, siguiendo con lo implementado por Mathieson & Mathieson, se descartaron las primeras 6,000 iteraciones.



**Figura 9.- Tiempo al Ancestro Común más Reciente (TMRCA) en generaciones de los modelos implementados con el software startMRCA.** En el eje X se muestra el número de iteraciones ( $n = 10,000$ ) y en el eje Y el TMRCA en generaciones de la variante rs9833097 inferido en cada iteración. Cada punto representa el TMRCA calculado para cada iteración. La línea azul representa el ajuste de regresión local del TMRCA en generaciones

El TMRCA estimado para cada una de las 4 variantes, fue la media de las últimas 4,000 iteraciones de cada análisis implementado para cada variante, obteniendo un TMRCA de 550.58 generaciones (IC 95% = 549.69 - 551.47) equivalente a 15,966.82 años (IC 95% = 15,941.01 - 15,992.63) para la variante rs9833097, para la variante rs13090265 fue de 551.78 generaciones (IC 95% = 550.89 - 552.68) igual a 16,001.62 años (IC 95% = 15,975.81 - 16,027.72), de la variante rs6782475 fue de 492.60 generaciones (IC 95% = 491.77 - 493.44) transformado en 14,285.40 años (IC 95% = 14,261.33 - 14,309.76) y de la variante rs7626560 fue de 484.97 generaciones (IC 95% = 484.21 - 485.73) aproximadamente 14,064.13 años (IC 95% = 14,042.09 - 14,086.17) (**Fig. 10**). Estos resultados concuerdan con los valores pequeños de los coeficientes de selección calculados anteriormente, sugiriendo que el proceso de selección natural en *PPARG* ocurrió aproximadamente entre 14,000 - 16,000 años atrás.



**Figura 10.- Medias del TMRCA en generaciones de las variantes con valores altos de selección.** BP: Antes del presente; SNV: Variante de un Solo Nucleótido

Al comparar la media de TMRCA obtenida de las 10 repeticiones para cada una de las cuatro variantes se encontraron diferencias significativas entre el TMRCA de la variante rs9833097 con el de la variante rs6782475 ( $p = 0.003$ ) y rs7626560 ( $p > 0.0001$ ) (**Tabla 5**). Lo anterior puede explicarse con base en los valores de LD observados entre las cuatro variantes. Las variantes rs9833097 y rs13090265 que no tuvieron diferencia significativa tuvieron mayores valores de LD ( $r^2 = 97$ ,  $p = 0.561$ ), al igual que ocurre entre la variante rs6782475 y rs7626560 ( $r^2 = 97$ ,  $p = 0.407$ ). Sin embargo, en los TMRCA en los que existieron diferencias significativas, se observó un menor LD entre las variantes utilizadas para la comparación, como es el caso de la variante rs9833097 con rs6782475 ( $r^2 = 60$ ,  $p = 0.003$ ) y con rs7626560 ( $r^2 = 58$ ,  $p > 0.0001$ ) (**Fig. 7**). Esto sugiere que el proceso de selección en *PPARG* para la población indígena se dio de forma gradual a lo largo del haplotipo y es consistente con el modelo de autostop génico, el cual pudo iniciar desde hace casi 16 mil años coincidente con la entrada del humano moderno a América.

**Tabla 5.- Comparación de las medias obtenidas por el bootstrapping (n=10) del TMRCA de las 4 variantes con valores más altos de selección.**

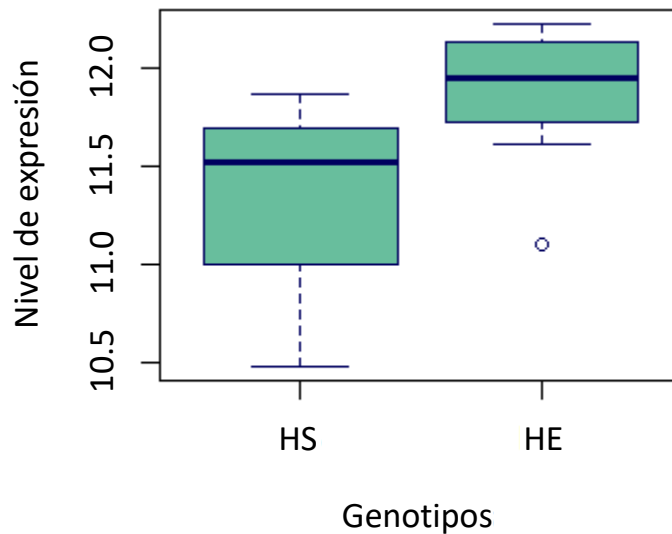
Variante media ± DE	rs13090265 550.51 ± 19.96	rs6782475 506.23 ± 20.78	rs7626560 492.57 ± 21.46	rs9833097 539.11 ± 14.79
rs13090265	-	> 0.0001*	> 0.0001*	0.561
rs6782475	> 0.0001*	-	0.407	0.003*
rs7626560	> 0.0001*	0.407	-	> 0.0001*
rs9833097	0.561	0.003*	> 0.0001*	-

\*:  $p < 0.05$  HSD Tukey; DE: Desviación estándar

#### 6.4 Análisis diferencial genómico

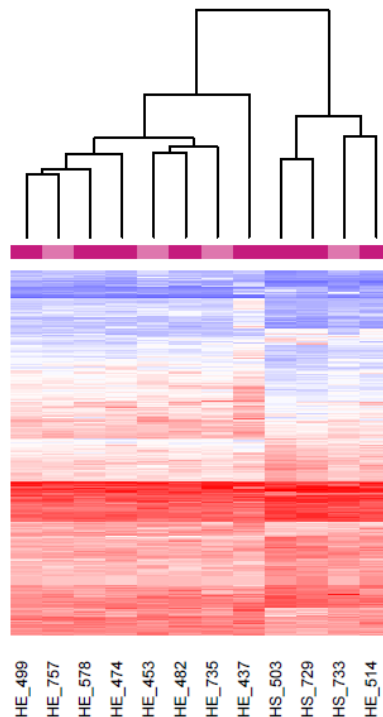
Con base en el análisis de las interacciones del promotor del gen *PPARG* con regiones enhancer alrededor de él, en el que observamos que algunas de estas variantes se encuentran en regiones reguladoras tipo “enhancer” o promotor que interaccionan con el promotor de *PPARG*, podemos sugerir que el haplotipo bajo selección puede estar modificando la expresión de este gen de forma diferencial entre portadores del haplotipo bajo selección vs portadores del haplotipo silvestre o de referencia.

Para comprobar lo anterior, se compararon los valores de los niveles de expresión de este gen entre los individuos mestizos HS (n = 3) y los individuos mestizos HE (n = 9). En un primer acercamiento, observamos una diferencia entre los niveles de expresión de *PPARG* entre ambos grupos, en la que los HE muestran niveles de expresión más elevados de *PPARG* en comparación con los individuos HS, sin embargo, estas diferencias no son significativas ( $p = 0.145$ ) (**Fig. 11**). Se espera que al incluir datos de expresión de individuos homocigotos para el haplotipo bajo selección en la comparación se puedan observar diferencias significativa en los resultados, por lo que es necesario hacer estudios posteriores para comprobar o rechazar esta hipótesis.



**Figura 11.- Niveles de expresión del gen *PPARG* entre portadores del haplotipo silvestre en estado homocigoto (HS,  $n = 3$ ) y heterocigotos que portan una copia del haplotipo bajo selección (HE,  $n = 9$ ).** La prueba de U de Mann-Whitney no mostró diferencias significativas entre ambos grupos ( $p = 0.145$ ).

Posteriormente, se realizó un análisis diferencial genómico, comparando los niveles de expresión de 27,189 transcritos entre HS vs HE, utilizando como punto de corte un  $FC > |1|$  y un valor de  $p < 0.05$  para definir a un transcrito diferencialmente expresado. Basado en lo anterior identificamos un total de 281 transcritos diferencialmente expresados entre HS y HE. El agrupamiento jerárquico no supervisado de estos 281 transcritos arrojó 2 grupos, el primero compuesto por los HS y el segundo por los HE, descartando la posibilidad de que este agrupamiento fuera debido al fenotipo diabético o no diabético de las muestras utilizadas en este análisis, ya que este fenotipo se distribuyó dentro de ambos grupos. Lo anterior sugiere que la presencia del haplotipo bajo selección de *PPARG* es capaz de generar una firma de expresión particular aún en individuos heterocigotos (**Fig. 12**).



**Figura 12.- Mapa de calor de los 281 transcritos diferencialmente expresados entre HS y HE.** Los colores de rojo a verde representan los niveles de expresión de los transcritos diferencialmente expresados ( $n = 281$  transcritos,  $FC > |1|$ ,  $p < 0.05$ ) entre los individuos homocigotos silvestres ( $n = 3$ ) y heterocigotos ( $n = 9$ ). El dendrograma representa el agrupamiento jerárquico de acuerdo con el genotipo (eje X). Las filas representan a los 281 transcritos (eje Y). Las barras en rosado claro debajo del dendrograma del eje X representan a los individuos no diabéticos y las barras rosadas oscuras a los diabéticos.

Dentro del Top 20 de los transcritos que tuvieron los valores más significativos, el 55% se encontraron sub-expresados (**Tabla 6**). Además, mediante una revisión en la plataforma [www.genecards.org](http://www.genecards.org) (Belinky et al., 2015) de las rutas y fenotipos a los que han sido asociados estos genes, se encontraron genes relevantes para el metabolismo de lípidos, como *LDLRAP1*, *NQO1*, *PPARA*, *SPTY2D1* y *BPNT2*; el sistema inmunológico, como *ZSWIM4*, *KLHL5*, *SSPN*, *CIR1*, *RNASEH2B* y *GOLGA8K*; la proliferación celular, como *ACTR1* y *MAGEH1*; y la homeostasis de la glucosa, como *GASK1B* (**Tabla 6**). También se revisaron los factores de transcripción que se unen a sitios Enhancer/Promotor a los que han sido asociados estos genes y se encontró que el 65% de ellos están asociados a al menos una región enhancer/promotor a la que se une *PPARG* como factor de transcripción (**Tabla 6**).



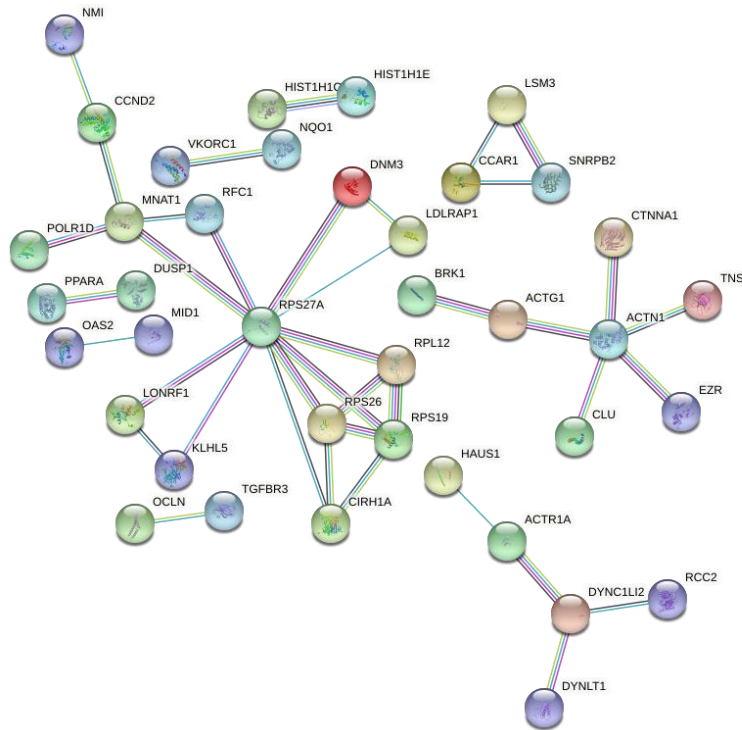
**Tabla 6.- Top 20 de los genes con los valores más significativos ( $FC > |1|$ ,  $P < 0.05$ ).**

#Chr	Gen	Log(FC)	p
19	ZSWIM4	1.30194563	2.05E-05
4	GASK1B	-1.46992574	2.96E-05
22	SMDT1	1.13273855	0.00012767
4	KLHL5*	-1.38679639	0.00019384
12	SSPN*	-1.20580056	0.00029418
15	GOLGA8M	1.33333224	0.00041623
10	ACTR1A*	-1.24508318	0.00048712
17	MYO1C*	-1.2167031	0.00050327
2	TMEM150A*	-1.06813147	0.00053403
1	LDLRAP1*	-1.03669443	0.00056223
19	SCAMP4*	-1.36974657	0.00060043
16	NQO1*	-1.50978323	0.00061032
2	CIR1*	1.08706977	0.00061793
22	PPARA*	-1.09011837	0.00062705
11	SPTY2D1*	1.07702497	0.00067982
X	MAGEH1	1.05933397	0.0007031
8	BPNT2*	-1.12019847	0.00082561
13	RNASEH2B	1.3234016	0.00083616
16	FLYWCH2*	1.09026699	0.00098948
15	GOLGA8K	1.42574002	0.0010065

\*: Genes asociados a al menos una región Promotor/Enhancer a la que se une *PPARG*

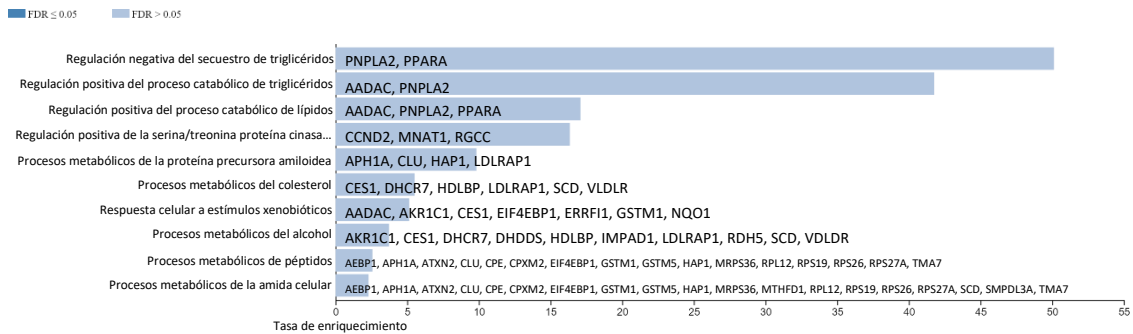
El análisis de red de interacción génica entre los 281 transcritos significativos reveló 9 redes de interacción que involucran por lo menos a dos transcritos (**Fig. 13**). Dentro de estas no se observó algún patrón en las funciones a las que se asocian los transcritos que integraron las redes. Sin embargo, nuevamente aparecieron transcritos relacionados al metabolismo de lípidos identificados dentro

del Top 20, como *LDLRAP1*, *NQO1* y *PPARA*; a la proliferación celular, como *ACTR1A*; y al sistema inmunológico, como *KLHL5* (**Fig. 13, Tabla 6**).



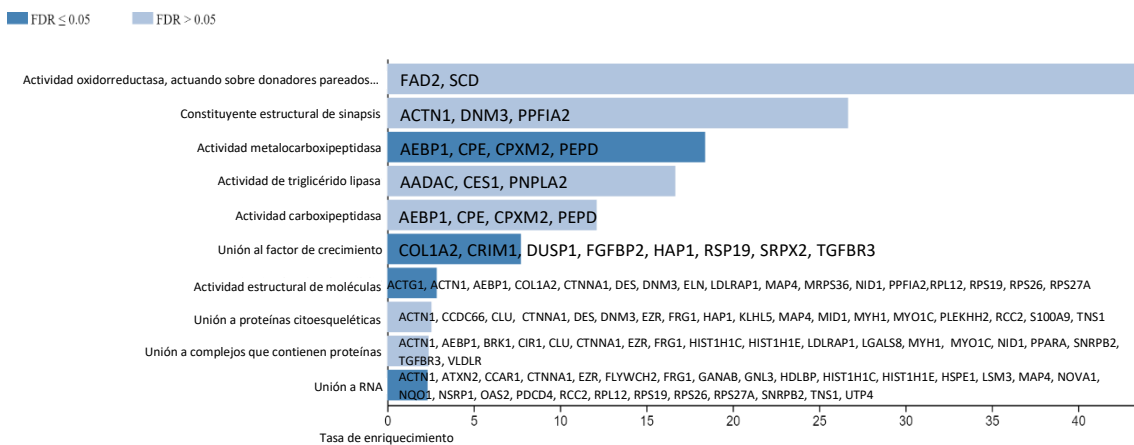
**Figura 13.- Red de interacciones de los genes expresados diferencialmente de manera significativa ( $n = 281$  genes,  $FC > |1|$ ,  $p < 0.05$ ) entre los individuos homocigotos silvestres y heterocigotos.** Se eliminaron los nodos sin conexión y con puntaje de interacción menor a 0.900

En los análisis de enriquecimiento, tomando en consideración el basado en procesos biológicos, observamos que a pesar de no haber agrupamientos significativos ( $FDR > 0.05$ , del inglés False Discovery Rate), uno de los procesos biológicos está relacionado a la proliferación celular (regulación positiva de la actividad de la proteína ciclina cinasa dependiente de serina/treonina) y cuatro se encuentran relacionados con el metabolismo de lípidos (regulación negativa del secuestro de triglicéridos, regulación positiva del proceso catabólico de los triglicéridos, regulación positiva del proceso catabólico de los lípidos y proceso metabólico del colesterol; **Fig. 14**).



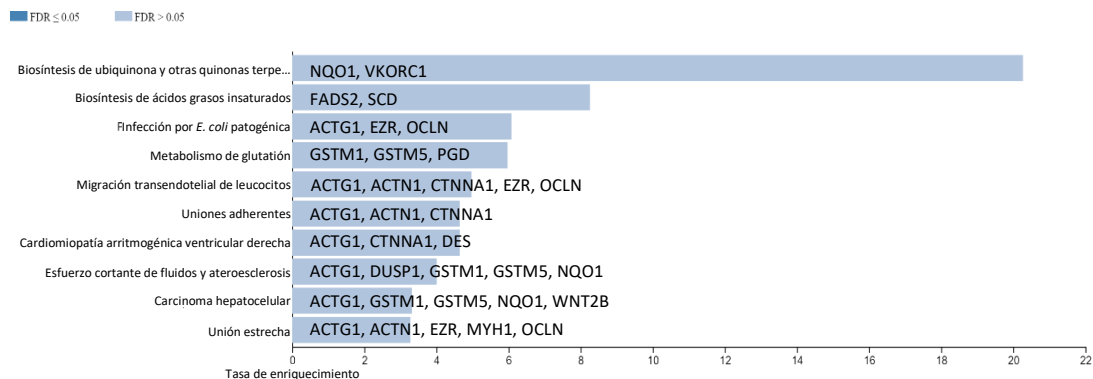
**Figura 14.- Análisis de enriquecimiento de los procesos biológicos en los que participan los genes expresados diferencialmente de manera significativa (n = 281 genes, FC > |1|, p < 0.05) entre los individuos HS y HE.**

El segundo agrupamiento realizado con base en las funciones moleculares descritas para cada gen, se observó el enriquecimiento de una función relacionada nuevamente al metabolismo de lípidos (actividad de la lipasa de triglicéridos). Además, en al menos la mitad de estas funciones tanto significativas (FDR < 0.05) como no significativas (FDR > 0.05), estuvieron enriquecidas por genes relacionados al metabolismo de los lípidos (*FADS2*, *SCD*, *AEBP1*, *CPE*, *LDLRAP1*, *HDLBP*; Fig. 15).



**Figura 15.- Análisis de enriquecimiento de las funciones moleculares en las que participan los genes expresados diferencialmente de manera significativa entre los individuos HS y HE (n = 281 genes, FC > |1|, p < 0.05).**

Por último, se realizó el análisis de enriquecimiento de rutas metabólicas utilizando la base de datos de Kegg. De nuevo se observó que al menos una de las rutas enriquecidas estuvo relacionada al metabolismo de lípidos (biosíntesis de ácidos grasos insaturados) aunque no de manera significativa (FDR > 0.05) (**Fig. 16**).



**Figura 16.- Análisis de enriquecimiento de las rutas metabólicas en las que participan los genes expresados diferencialmente de manera significativa ( $n = 281$  genes,  $FC > |1|$ ,  $p < 0.05$ ) entre los individuos HS y HE.**

## VII. Discusión

El estudio de la dinámica evolutiva permite generar mejores modelos de la diversidad genética regional, lo cual es importante para el desarrollo y la aplicación de la medicina personalizada en las poblaciones (O'Connor, 2018). Analizar las enfermedades desde el punto de vista de la evolución, brinda una poderosa referencia para analizar la interacción que se da entre el genotipo que subyace a una enfermedad y un ambiente específico. Por ejemplo, algunas enfermedades resultan de la interacción entre las variantes genéticas de una población específica con cambios rápidos y drásticos en el ambiente en el que se desarrolló la población. Tal es el caso de la obesidad que, de acuerdo a la hipótesis del genotipo ahorrador, podría ser el resultado de una mala interacción entre las variantes que modifican la actividad de los genes que subyacen a este padecimiento y los cambios drásticos en sus estilos de vida, como un consumo cada vez más elevado de alimentos altos en calorías, mayor sedentarismo y menor cantidad de horas de sueño (Benton et al., 2021).

En este proyecto se identificó la estructura y posible funcionalidad del haplotipo con las variantes con señales de selección en el gen *PPARG* identificado en población indígena de México, el tiempo en que comenzó a incrementar su frecuencia y el fenotipo al que podría asociarse dicho haplotipo. Con respecto a la estructura del haplotipo bajo selección, se identificaron un total de 34 variantes en alto LD con la variante rs9833097 (**Tabla 1**). La comparación de este haplotipo con el de poblaciones africanas, europeas y asiáticas reveló que su estructura es específica de la población mexicana (**Fig. 7, Anexo 1-5**). Lo anterior también fue apoyado por la similitud del haplotipo identificado en este proyecto con el de la población mexicanoamericana MXL del proyecto 1000 Genomas (**Anexo 5a**).

Más aún, el análisis *in silico* en conjunto con el de las interacciones promotor-enhancer en *PPARG* demostraron que 10 de los SNVs del haplotipo bajo selección se encontraron en regiones reguladoras tipo enhancer/promotor que podrían ser importantes para la regulación transcripcional de este gen (**Tabla 3, Fig. 8**). En apoyo a esta aseveración, un estudio previo realizado en población indígena Pima,

demonstró que variantes en la región promotora de *PPARG*, podrían modificar la expresión del gen y conllevar a consecuencias en las rutas reguladas por *PPARG* (Muller et al., 2003). También se ha demostrado, mediante estudios en modelos animales, que las variantes con señales de selección en regiones intrónicas de un gen pueden estar asociadas a cambios en la regulación epigenética de una amplia región dentro de un locus llevando a cambios importantes en el fenotipo metabólico. Tal fue el caso del locus 2q21.3 que abarca variantes con señales de selección en genes como *LCT* y *R3HDM1* identificadas en población europea (L. Wang et al., 2020). Además, estudios *in vitro* han evidenciado la importancia de la desmetilación del promotor de *PPARG* durante la adipogénesis para promover que el locus de este gen sea accesible para factores de transcripción que incluyen al mismo *PPARG* (Fujiki et al., 2009; Wakabayashi et al., 2009). Por lo tanto, nuestros resultados sugieren que la selección de las variantes en *PPARG* en poblaciones indígenas de México condujo a la presencia de un haplotipo con estructura única en esta población, la cual puede estar relacionada con la regulación génica de *PPARG* y probablemente de sus genes blanco.

La comparación del coeficiente de selección del haplotipo en *PPARG* sugiere que el proceso de selección observado en este gen fue más débil ( $s = 0.0008 - 0.0042$ , **Tabla 4**) que el de otros eventos similares reportados para poblaciones humanas, como el de la región *LCT* en poblaciones europeas, el del gen *G6PD* en poblaciones africanas y en la región *MHC* en población mexicana ( $s = 0.09 - 0.19$ ,  $s = 0.02 - 0.05$  y  $s = 0.05 - 0.07$  respectivamente) (Zhou et al., 2016; Bersaglieri et al., 2004; Tishkoff, 2001).

Sin embargo, el tiempo que tarda una variante bajo selección en aumentar su frecuencia alélica es inversamente proporcional a la intensidad del proceso de selección bajo el cual ha estado, es decir, un mayor proceso de selección indica que el tiempo en el cual se seleccionó un locus es más reciente al presente y viceversa (Hamilton, 2009). Lo anterior puede ser ejemplificado con el caso de las variantes en la región *LCT* en poblaciones europeas, las cuales se ha estimado que comenzaron a incrementar su frecuencia aproximadamente 2,000 años antes del

presente (BP) ( $s = 0.09 - 0.19$ ), el gen de *FADS1* que se seleccionó aproximadamente hace 4000 años BP ( $s = 0.034 - 0.069$ ) o el gen *MHC* en mestizos mexicanos que se sugiere se seleccionó hace 500 años BP ( $s = 0.05 - 0.07$ ) (Bersaglieri et al., 2004; Mathieson & Mathieson, 2018; Zhou et al., 2016).

Con base en estos ejemplos, el  $s$  obtenido para el gen *PPARG* en población indígena mexicana puede sugerir un evento de selección aún más antiguo. En este mismo sentido, nuestros resultados del TMRCA de *PPARG* en población mexicana sugieren que el evento de selección comenzó hace ~15,000 años BP (**Fig. 10**), lo cual concuerda con el bajo coeficiente de selección que se observó. Notablemente, el TMRCA calculado en este estudio coincide con el tiempo en que se estima que el hombre moderno llegó al continente americano (16,000 a 13,500 años atrás) (Halligan et al., 2016; Fagundes et al., 2018; Llamas et al., 2016; Potter et al., 2018). Además, estos años marcan el término del último periodo glacial máximo durante el último milenio del pleistoceno tardío. Estudios previos, han sugerido que los primeros pobladores de América se enfrentaron a condiciones poco favorables previo a su llegada al continente como temperaturas invernales extremas, largos recorridos a pie y poca disponibilidad de alimentos (Hoffecker et al., 2016). Nuestros resultados en el TMRCA, junto a la información anterior, indican que estas condiciones adversas posiblemente actuaron como presiones selectivas sobre el haplotipo identificado bajo selección en población indígena mexicana. En este punto, cabe destacar que el gen *PPARG* tiene un papel importante en la regulación génica de la adipogénesis, la respuesta corporal a la insulina y la homeostasis de los lípidos y la glucosa (Janani & Ranjitha Kumari, 2015). Por lo que el haplotipo bajo selección en *PPARG* pudo haber contribuido de alguna manera a un manejo más eficaz de la energía en el organismo de los primeros pobladores americanos, lo que les otorgó a los portadores del haplotipo mayores probabilidades de sobrevivir a dicho evento de migración.

Dada la función crucial de *PPARG* en la regulación génica de distintas rutas metabólicas energéticas, nos dimos a la tarea de investigar una posible firma de selección mediante un análisis de expresión diferencial en la búsqueda de firmas de

expresión entre los individuos que portan estas variantes y los que no. En un primer acercamiento observamos que los niveles de expresión del gen *PPARG* no muestran diferencias estadísticamente significativas entre el grupo de los HE y los HS. Sin embargo, los HE muestran niveles mayores de expresión de *PPARG* (**Fig. 11**). Esta ausencia de significancia puede ser consecuencia de varios factores, entre ellos el tamaño de muestra utilizado para las comparaciones, ya que solo se incluyeron 3 homocigotos silvestres y 9 individuos heterocigotos, lo cual sugiere que nuestra muestra no tiene el suficiente poder estadístico (Krzywinski & Altman, 2013). Otro factor, es la limitación para acceder a individuos homocigotos para el haplotipo bajo selección en *PPARG*. Por lo tanto, esperamos que al aumentar el número de individuos homocigotos silvestres e incluir individuos homocigotos para el haplotipo bajo selección, se logre observar un aumento significativo en los niveles de expresión entre los diferentes grupos. Sin embargo, estos resultados preliminares, apuntan a que las variantes de *PPARG* bajo selección tienden a aumentar los niveles de expresión de *PPARG*, lo cual puede ser explicado por el hecho de que 10 de las variantes que componen este haplotipo se encuentran distribuidas en regiones tipo enhancer/promotor (**Fig. 8**). Sin embargo, se necesitan estudios adicionales para dilucidar el papel que juegan estas variantes sobre la expresión de *PPARG*.

Adicionalmente, al analizar el estado transcripcional global entre HE y HS, se encontraron 281 genes diferencialmente expresados entre ambos grupos. De forma interesante, dentro de estos genes, se encuentran algunos genes subexpresados (*LDLRAP1*, *NQO1*, *PPARA* y *BPNT2*) y sobreexpresados (*SPTY2D1*) en el grupo portador del haplotipo bajo selección que se han asociado a alguna característica del metabolismo de lípidos (**Tabla 6**). Respecto a los genes subexpresados, se ha identificado que en el gen *NQO1* existen variantes asociadas a fenotipos como un aumento de los niveles de triglicéridos en sangre y en el gen *LDLRAP1* variantes asociadas a una disminución de colesterol total y colesterol LDL (Buniello et al., 2019). Además, el gen *PPARA* es un gen que suele expresarse más en el tejido adiposo marrón que en el tejido adiposo blanco que es donde tiene mayor expresión *PPARG* (Saely et al., 2012). Por lo tanto, nuestros resultados sugieren que los



portadores del haplotipo bajo selección podrían tener una actividad metabólica más eficiente para el almacenamiento de lípidos y adipogénesis dirigida a la producción de tejido adiposo blanco. Ambas características podrían estar derivando en un fenotipo con mejor manejo de energía. Lo anterior es apoyado por nuestros resultados del análisis de enriquecimiento de procesos biológicos, funciones moleculares y rutas metabólicas de los 281 genes expresados diferencialmente, en donde se agruparon distintos genes relacionados al metabolismo de lípidos (**Fig. 14-16**). Lo cual tiene sentido si tomamos en cuenta que las variantes bajo selección podrían estar causando modificaciones en la regulación de la expresión de un gen que a su vez regula rutas metabólicas importantes para el mantenimiento de la homeostasis de lípidos.

## VIII. Conclusiones

- Las variantes que se encuentran bajo selección positiva en el gen *PPARG* parecen no ejercer un efecto directo sobre la secuencia o estructura de la proteína que codifican, más bien sugieren que están participando en la regulación a nivel de expresión, sin embargo, se requieren de más estudios para comprobarlo.
- Nuestras inferencias acerca del TMRCA para el gen *PPARG*, sugieren que el haplotipo bajo selección comenzó a incrementar su frecuencia durante el poblamiento inicial de América.
- Los datos sugieren que los portadores del haplotipo bajo selección en *PPARG* muestran una firma de expresión particular en tejido adiposo subcutáneo en vías relacionadas con el metabolismo de los lípidos.
- El haplotipo bajo selección pudo representar una ventaja que favoreció la migración de los primeros pobladores al continente americano al participar en un manejo eficaz de la energía en el organismo.
- Hasta donde es de nuestro conocimiento, este trabajo representa la primera evidencia de como haplotipos bajo selección natural en *PPARG* en población indígena mexicana influyen sobre la aparición de una firma de expresión específica, relacionada al metabolismo y biogénesis de las grasas.

## IX. Referencias bibliográficas

- Artasensi, A., Pedretti, A., Vistoli, G., & Fumagalli, L. (2020). Type 2 Diabetes Mellitus: A Review of Multi-Target Drugs. *Molecules*, 25(8), 1987. <https://doi.org/10.3390/molecules25081987>
- Auton, A., & McVean, G. (2012). Estimating Recombination Rates from Genetic Variation in Humans. En M. Anisimova (Ed.), *Evolutionary Genomics* (Vol. 856, pp. 217-237). Humana Press. [https://doi.org/10.1007/978-1-61779-585-5\\_9](https://doi.org/10.1007/978-1-61779-585-5_9)
- Barajas-Olmos, F., Centeno-Cruz, F., Zerrweck, C., Imaz-Rosshandler, I., Martínez-Hernández, A., Cordova, E. J., Rangel-Escareño, C., Gálvez, F., Castillo, A., Maydón, H., Campos, F., Maldonado-Pintado, D. G., & Orozco, L. (2018). Altered DNA methylation in liver and adipose tissues derived from individuals with obesity and type 2 diabetes. *BMC Medical Genetics*, 19(1), 28. <https://doi.org/10.1186/s12881-018-0542-8>
- Belinky, F., Nativ, N., Stelzer, G., Zimmerman, S., Iny Stein, T., Safran, M., & Lancet, D. (2015). PathCards: Multi-source consolidation of human biological pathways. *Database*, 2015. <https://doi.org/10.1093/database/bav006>
- Benton, M. L., Abraham, A., LaBella, A. L., Abbot, P., Rokas, A., & Capra, J. A. (2021). The influence of evolutionary history on human health and disease. *Nature Reviews Genetics*, 22(5), 269-283. <https://doi.org/10.1038/s41576-020-00305-9>
- Bersaglieri, T., Sabeti, P. C., Patterson, N., Vanderploeg, T., Schaffner, S. F., Drake, J. A., Rhodes, M., Reich, D. E., & Hirschhorn, J. N. (2004). Genetic

Signatures of Strong Recent Positive Selection at the Lactase Gene.

*American Journal of Human Genetics*, 74(6), 1111-1120.

Broekema, M. F., Savage, D. B., Monajemi, H., & Kalkhoven, E. (2019). Gene-gene and gene-environment interactions in lipodystrophy: Lessons learned from natural PPAR $\gamma$  mutants. *Biochimica et Biophysica Acta (BBA) - Molecular and Cell Biology of Lipids*, 1864(5), 715-732.  
<https://doi.org/10.1016/j.bbailip.2019.02.002>

Buniello, A., MacArthur, J. A. L., Cerezo, M., Harris, L. W., Hayhurst, J., Malangone, C., McMahon, A., Morales, J., Mountjoy, E., Sollis, E., Suveges, D., Vrousseau, O., Whetzel, P. L., Amode, R., Guillen, J. A., Riat, H. S., Trevanion, S. J., Hall, P., Junkins, H., ... Parkinson, H. (2019). The NHGRI-EBI GWAS Catalog of published genome-wide association studies, targeted arrays and summary statistics 2019. *Nucleic Acids Research*, 47(D1), D1005-D1012. <https://doi.org/10.1093/nar/gky1120>

Capaccio, D., Ciccodicola, A., Sabatino, L., Casamassimi, A., Pancione, M., Fucci, A., Febbraro, A., Merlino, A., Graziano, G., & Colantuoni, V. (2010). A novel germline mutation in Peroxisome Proliferator-Activated Receptor  $\gamma$  gene associated with large intestine polyp formation and dyslipidemia. *Biochimica et Biophysica Acta (BBA) - Molecular Basis of Disease*, 1802(6), 572-581.  
<https://doi.org/10.1016/j.bbadis.2010.01.012>

Casillas, S., & Barbadilla, A. (2017). Molecular Population Genetics. *Genetics*, 205(3), 1003-1035. <https://doi.org/10.1534/genetics.116.196493>

Costa, V., Gallo, M. A., Letizia, F., Aprile, M., Casamassimi, A., & Ciccodicola, A. (2010). PPAR $\gamma$ : Gene Expression Regulation and Next-Generation

Sequencing for Unsolved Issues. *PPAR Research*, 2010, 1-17.

<https://doi.org/10.1155/2010/409168>

Damgaard, P. de B., Marchi, N., Rasmussen, S., Peyrot, M., Renaud, G., Korneliussen, T., Moreno-Mayar, J. V., Pedersen, M. W., Goldberg, A., Usmanova, E., Baimukhanov, N., Loman, V., Hedeager, L., Pedersen, A. G., Nielsen, K., Afanasiev, G., Akmatov, K., Aldashev, A., Alpaslan, A., ... Willerslev, E. (2018). 137 ancient human genomes from across the Eurasian steppes. *Nature*, 557(7705), 369-374. <https://doi.org/10.1038/s41586-018-0094-2>

Duarte, C. W., Vaughan, L. K., Mark Beasley, T., & Tiwari, H. K. (2014).

Multifactorial Inheritance and Complex Diseases. En *Reference Module in Biomedical Sciences*. Elsevier. <https://doi.org/10.1016/B978-0-12-801238-3.05485-4>

Ellegren, H., & Galtier, N. (2016). Determinants of genetic diversity. *Nature Reviews Genetics*, 17(7), 422-433. <https://doi.org/10.1038/nrg.2016.58>

Ewbank, D. (2016). Measuring selection in human populations using the growth rate per generation. *Philosophical Transactions of the Royal Society B: Biological Sciences*, 371(1692), 20150148.

<https://doi.org/10.1098/rstb.2015.0148>

Fagundes, N. J. R., Tagliani-Ribeiro, A., Rubicz, R., Tarskaia, L., Crawford, M. H., Salzano, F. M., & Bonatto, S. L. (2018). How strong was the bottleneck associated to the peopling of the Americas? New insights from multilocus sequence data. *Genetics and Molecular Biology*, 41(1 suppl 1), 206-214. <https://doi.org/10.1590/1678-4685-gmb-2017-0087>

- Fan, S., Hansen, M. E. B., Lo, Y., & Tishkoff, S. A. (2016). Going global by adapting local: A review of recent human adaptation. *Science*, 354(6308), 54-59. <https://doi.org/10.1126/science.aaf5098>
- Fujiki, K., Kano, F., Shiota, K., & Murata, M. (2009). Expression of the peroxisome proliferator activated receptor gamma gene is repressed by DNA methylation in visceral adipose tissue of mouse models of diabetes. *BMC Biology*, 7(1), 38. <https://doi.org/10.1186/1741-7007-7-38>
- García-Ortiz, H., Barajas-Olmos, F., Contreras-Cubas, C., Cid-Soto, M. Á., Cordova, E. J., Centeno-Cruz, F., Mendoza-Caamal, E., Cicerón-Arellano, I., Flores-Huacuja, M., Baca, P., Bolnick, D. A., Snow, M., Flores-Martínez, S. E., Ortíz-López, R., Reynolds, A. W., Morales-Marín, M., Velázquez-Cruz, R., Kostic, A. D., Galaviz-Hernández, C., ... Orozco, L. (2021). The genomic landscape of Mexican Indigenous populations brings insights into the peopling of the Americas. *Nature Communications*.
- Halligan, J. J., Waters, M. R., Perrotti, A., Owens, I. J., Feinberg, J. M., Bourne, M. D., Fenerty, B., Winsborough, B., Carlson, D., Fisher, D. C., Stafford, T. W., & Dunbar, J. S. (2016). Pre-Clovis occupation 14,550 years ago at the Page-Ladson site, Florida, and the peopling of the Americas. *Science Advances*, 2(5), e1600375. <https://doi.org/10.1126/sciadv.1600375>
- Hamilton, M. B. (2009). *Population genetics*. Wiley-Blackwell.
- Harris, D. N., Song, W., Shetty, A. C., Levano, K. S., Cáceres, O., Padilla, C., Borda, V., Tarazona, D., Trujillo, O., Sanchez, C., Kessler, M. D., Galarza, M., Capristano, S., Montejo, H., Flores-Villanueva, P. O., Tarazona-Santos, E., O'Connor, T. D., & Guio, H. (2018). Evolutionary genomic dynamics of

- Peruvians before, during, and after the Inca Empire. *Proceedings of the National Academy of Sciences*, 115(28), E6526-E6535.  
<https://doi.org/10.1073/pnas.1720798115>
- Hartl, D. L., & Clark, A. G. (2007). *Principles of population genetics* (4th ed). Sinauer Associates.
- Hedrick, P. W. (2011). *Genetics of populations* (4th ed). Jones and Bartlett Publishers.
- Heikkinen, S., Auwerx, J., & Argmann, C. A. (2007). PPAR $\gamma$  in human and mouse physiology. *Biochimica et biophysica acta*, 1771(8), 999-1013.  
<https://doi.org/10.1016/j.bbaliip.2007.03.006>
- Hoffecker, J. F., Elias, S. A., O'Rourke, D. H., Scott, G. R., & Bigelow, N. H. (2016). Beringia and the global dispersal of modern humans: Beringia and the Global Dispersal of Modern Humans. *Evolutionary Anthropology: Issues, News, and Reviews*, 25(2), 64-78. <https://doi.org/10.1002/evan.21478>
- Janani, C., & Ranjitha Kumari, B. D. (2015). PPAR gamma gene – A review. *Diabetes & Metabolic Syndrome: Clinical Research & Reviews*, 9(1), 46-50.  
<https://doi.org/10.1016/j.dsx.2014.09.015>
- Joffe, Y. T., & Houghton, C. A. (2016). A Novel Approach to the Nutrigenetics and Nutrigenomics of Obesity and Weight Management. *Current Oncology Reports*, 18(7), 43. <https://doi.org/10.1007/s11912-016-0529-6>
- Kanehisa, M., Furumichi, M., Sato, Y., Ishiguro-Watanabe, M., & Tanabe, M. (2021). KEGG: Integrating viruses and cellular organisms. *Nucleic Acids Research*, 49(D1), D545-D551. <https://doi.org/10.1093/nar/gkaa970>

- Keith, J. M. (Ed.). (2008). *Bioinformatics* (Vol. 452). Humana Press.  
<https://doi.org/10.1007/978-1-60327-159-2>
- Kern, A. D., & Hahn, M. W. (2018). The Neutral Theory in Light of Natural Selection. *Molecular Biology and Evolution*, *35*(6), 1366-1371.  
<https://doi.org/10.1093/molbev/msy092>
- Krzywinski, M., & Altman, N. (2013). Power and sample size. *Nature Methods*, *10*(12), 1139-1140. <https://doi.org/10.1038/nmeth.2738>
- Lazaridis, I., Patterson, N., Mitnik, A., Renaud, G., Mallick, S., Kirsanow, K., Sudmant, P. H., Schraiber, J. G., Castellano, S., Lipson, M., Berger, B., Economou, C., Bollongino, R., Fu, Q., Bos, K. I., Nordenfelt, S., Li, H., de Filippo, C., Prüfer, K., ... Krause, J. (2014). Ancient human genomes suggest three ancestral populations for present-day Europeans. *Nature*, *513*(7518), 409-413. <https://doi.org/10.1038/nature13673>
- Lin, M., Siford, R. L., Martin, A. R., Nakagome, S., Möller, M., Hoal, E. G., Bustamante, C. D., Gignoux, C. R., & Henn, B. M. (2018). Rapid evolution of a skin-lightening allele in southern African KhoeSan. *Proceedings of the National Academy of Sciences*, *115*(52), 13324-13329.  
<https://doi.org/10.1073/pnas.1801948115>
- Lindo, J., Huerta-Sánchez, E., Nakagome, S., Rasmussen, M., Petzelt, B., Mitchell, J., Cybulski, J. S., Willerslev, E., DeGiorgio, M., & Malhi, R. S. (2016). A time transect of exomes from a Native American population before and after European contact. *Nature Communications*, *7*(1), 13175.  
<https://doi.org/10.1038/ncomms13175>



- Liu, Y., Wang, J., Luo, S., Zhan, Y., & Lu, Q. (2020). The roles of PPAR $\gamma$  and its agonists in autoimmune diseases: A comprehensive review. *Journal of Autoimmunity*, 113, 102510. <https://doi.org/10.1016/j.jaut.2020.102510>
- Llamas, B., Fehren-Schmitz, L., Valverde, G., Soubrier, J., Mallick, S., Rohland, N., Nordenfelt, S., Valdiosera, C., Richards, S. M., Rohrlach, A., Romero, M. I. B., Espinoza, I. F., Cagigao, E. T., Jiménez, L. W., Makowski, K., Reyna, I. S. L., Lory, J. M., Torrez, J. A. B., Rivera, M. A., ... Haak, W. (2016). Ancient mitochondrial DNA provides high-resolution time scale of the peopling of the Americas. *Science Advances*, 2(4), e1501385. <https://doi.org/10.1126/sciadv.1501385>
- Majithia, A. R., Flannick, J., Shahinian, P., Guo, M., Bray, M.-A., Fontanillas, P., Gabriel, S. B., GoT2D Consortium, NHGRI JHS/FHS Allelic Spectrum Project, SIGMA T2D Consortium, T2D-GENES Consortium, Rosen, E. D., Altshuler, D., Flannick, J., Manning, A. K., Hartl, C., Agarwala, V., Fontanillas, P., Green, T., ... Zollner, S. (2014). Rare variants in PPARG with decreased activity in adipocyte differentiation are associated with increased risk of type 2 diabetes. *Proceedings of the National Academy of Sciences*, 111(36), 13127-13132. <https://doi.org/10.1073/pnas.1410428111>
- Masel, J. (2011). Genetic drift. *Current Biology*, 21(20), R837-R838. <https://doi.org/10.1016/j.cub.2011.08.007>
- Mathieson, S., & Mathieson, I. (2018). FADS1 and the Timing of Human Adaptation to Agriculture. *Molecular Biology and Evolution*, 35(12), 2957-2970. <https://doi.org/10.1093/molbev/msy180>

- Mendoza-Caamal, E. C., Barajas-Olmos, F., García-Ortiz, H., Cicerón-Arellano, I., Martínez-Hernández, A., Córdova, E. J., Esparza-Aguilar, M., Contreras-Cubas, C., Centeno-Cruz, F., Cid-Soto, M., Morales-Marín, M. E., Reséndiz-Rodríguez, A., Jiménez-Ruiz, J. L., Salas-Martínez, M. G., Saldaña-Alvarez, Y., Mirzaeicheshmeh, E., Rojas-Martínez, M. R., & Orozco, L. (2020). Metabolic syndrome in indigenous communities in Mexico: A descriptive and cross-sectional study. *BMC Public Health*, *20*(1), 339.  
<https://doi.org/10.1186/s12889-020-8378-5>
- Mi, H., Muruganujan, A., Ebert, D., Huang, X., & Thomas, P. D. (2019). PANTHER version 14: More genomes, a new PANTHER GO-slim and improvements in enrichment analysis tools. *Nucleic Acids Research*, *47*(D1), D419-D426.  
<https://doi.org/10.1093/nar/gky1038>
- Muller, Y. L., Bogardus, C., Beamer, B. A., Shuldiner, A. R., & Baier, L. J. (2003). A Functional Variant in the Peroxisome Proliferator-Activated Receptor 2 Promoter Is Associated With Predictors of Obesity and Type 2 Diabetes in Pima Indians. *Diabetes*, *52*(7), 1864-1871.  
<https://doi.org/10.2337/diabetes.52.7.1864>
- O'Connor, T. D. (2018). Native American Genomic Diversity through Ancient DNA. *Cell*, *175*(5), 1173-1174. <https://doi.org/10.1016/j.cell.2018.10.058>
- Perry, G. H. (2008). The evolutionary significance of copy number variation in the human genome. *Cytogenetic and Genome Research*, *123*(1-4), 283-287.  
<https://doi.org/10.1159/000184719>
- Potter, B. A., Baichtal, J. F., Beaudoin, A. B., Fehren-Schmitz, L., Haynes, C. V., Holliday, V. T., Holmes, C. E., Ives, J. W., Kelly, R. L., Llamas, B., Malhi, R.

- S., Miller, D. S., Reich, D., Reuther, J. D., Schiffels, S., & Surovell, T. A. (2018). Current evidence allows multiple models for the peopling of the Americas. *Science Advances*, 4(8), eaat5473.  
<https://doi.org/10.1126/sciadv.aat5473>
- Powell, E., Kuhn, P., & Xu, W. (2007). Nuclear Receptor Cofactors in PPAR  $\gamma$  - Mediated Adipogenesis and Adipocyte Energy Metabolism. *PPAR Research*, 2007, 1-11. <https://doi.org/10.1155/2007/53843>
- Quinn, J. P., Savage, A. L., & Bubb, V. J. (2019). Non-coding genetic variation shaping mental health. *Current Opinion in Psychology*, 27, 18-24.  
<https://doi.org/10.1016/j.copsyc.2018.07.006>
- Raghavan, M., Steinrücken, M., Harris, K., Schiffels, S., Rasmussen, S., DeGiorgio, M., Albrechtsen, A., Valdiosera, C., Ávila-Arcos, M. C., Malaspina, A.-S., Eriksson, A., Moltke, I., Metspalu, M., Homburger, J. R., Wall, J., Cornejo, O. E., Moreno-Mayar, J. V., Korneliussen, T. S., Pierre, T., ... Willerslev, E. (2015). Genomic evidence for the Pleistocene and recent population history of Native Americans. *Science*, 349(6250), aab3884.  
<https://doi.org/10.1126/science.aab3884>
- Rees, J. S., Castellano, S., & Andrés, A. M. (2020). The Genomics of Human Local Adaptation. *Trends in Genetics*, 36(6), 415-428.  
<https://doi.org/10.1016/j.tig.2020.03.006>
- Reynolds, A. W., Mata-Míguez, J., Miró-Herrans, A., Briggs-Cloud, M., Sylestine, A., Barajas-Olmos, F., Garcia-Ortiz, H., Rzhetskaya, M., Orozco, L., Raff, J. A., Hayes, M. G., & Bolnick, D. A. (2019). Comparing signals of natural selection between three Indigenous North American populations.

*Proceedings of the National Academy of Sciences*, 116(19), 9312-9317.

<https://doi.org/10.1073/pnas.1819467116>

Ristow, M., Müller-Wieland, D., Pfeiffer, A., Krone, W., & Kahn, C. R. (1998).

Obesity Associated with a Mutation in a Genetic Regulator of Adipocyte Differentiation. *New England Journal of Medicine*, 339(14), 953-959.

<https://doi.org/10.1056/NEJM199810013391403>

Rohde, K., Keller, M., la Cour Poulsen, L., Blüher, M., Kovacs, P., & Böttcher, Y.

(2019). Genetics and epigenetics in obesity. *Metabolism*, 92, 37-50.

<https://doi.org/10.1016/j.metabol.2018.10.007>

Romero, F., & Farias, J. (2014). *La medicina evolutiva o darwiniana*. 57, 5-14.

Sabeti, P. C., Schaffner, S. F., Fry, B., Lohmueller, J., Varilly, P., Shamovsky, O.,

Palma, A., Mikkelsen, T. S., Altshuler, D., & Lander, E. S. (2006). Positive Natural Selection in the Human Lineage. *Science*, 312(5780), 1614-1620.

<https://doi.org/10.1126/science.1124309>

Saeb, A. T. M., & Al-Naqeb, D. (2016). The Impact of Evolutionary Driving Forces on Human Complex Diseases: A Population Genetics Approach. *Scientifica*,

2016, 1-10. <https://doi.org/10.1155/2016/2079704>

Saely, C. H., Geiger, K., & Drexel, H. (2012). Brown versus White Adipose Tissue: A Mini-Review. *Gerontology*, 58(1), 15-23.

<https://doi.org/10.1159/000321319>

Secretaria de Salud. (2005). *Desafíos para el sistema mexicano de salud*.

<http://www.salud.gob.mx/unidades/evaluacion/saludmex2005/SM-2001-05cap3.pdf>

- Shamah-Levy, T., Vielma-Orozco, E., Heredia-Hernández, O., Romero-Martínez, M., Mojica-Cuevas, J., Cuevas-Nasu, L., Santaella-Castell, J., & Rivera-Dommarco, J. (2020). *Encuesta Nacional de Salud y Nutrición 2018-19. Resultados nacionales*. Instituto Nacional de Salud Pública.  
[https://ensanut.insp.mx/encuestas/ensanut2018/doctos/informes/ensanut\\_2018\\_informe\\_final.pdf](https://ensanut.insp.mx/encuestas/ensanut2018/doctos/informes/ensanut_2018_informe_final.pdf)
- Slatkin, M. (2008). Linkage disequilibrium—Understanding the evolutionary past and mapping the medical future. *Nature Reviews Genetics*, 9(6), 477-485.  
<https://doi.org/10.1038/nrg2361>
- Smith, J., Coop, G., Stephens, M., & Novembre, J. (2018). Estimating Time to the Common Ancestor for a Beneficial Allele. *Molecular Biology and Evolution*, 35(4), 1003-1017. <https://doi.org/10.1093/molbev/msy006>
- Stern, A. J., & Nielsen, R. (2019). Detecting Natural Selection. En D. Balding, I. Moltke, & J. Marioni (Eds.), *Handbook of Statistical Genomics* (1.<sup>a</sup> ed., pp. 397-40). Wiley. <https://doi.org/10.1002/9781119487845.ch14>
- Sved, J. A., & Hill, W. G. (2018). One Hundred Years of Linkage Disequilibrium. *Genetics*, 209(3), 629-636. <https://doi.org/10.1534/genetics.118.300642>
- The 1000 Genomes Project Consortium. (2015). A global reference for human genetic variation. *Nature*, 526(7571), 68-74.  
<https://doi.org/10.1038/nature15393>
- The ENCODE Project Consortium, Moore, J. E., Purcaro, M. J., Pratt, H. E., Epstein, C. B., Shores, N., Adrian, J., Kawli, T., Davis, C. A., Dobin, A., Kaul, R., Halow, J., Van Nostrand, E. L., Freese, P., Gorkin, D. U., Shen, Y., He, Y., Mackiewicz, M., Pauli-Behn, F., ... Weng, Z. (2020). Expanded

- encyclopaedias of DNA elements in the human and mouse genomes. *Nature*, 583(7818), 699-710. <https://doi.org/10.1038/s41586-020-2493-4>
- Thurman, T. J., & Barrett, R. D. H. (2016). The genetic consequences of selection in natural populations. *Molecular Ecology*, 25(7), 1429-1448. <https://doi.org/10.1111/mec.13559>
- Tiffin, P., & Ross-Ibarra, J. (2014). Advances and limits of using population genetics to understand local adaptation. *Trends in Ecology & Evolution*, 29(12), 673-680. <https://doi.org/10.1016/j.tree.2014.10.004>
- Tigano, A., & Friesen, V. L. (2016). Genomics of local adaptation with gene flow. *Molecular Ecology*, 25(10), 2144-2164. <https://doi.org/10.1111/mec.13606>
- Tishkoff, S. A. (2001). Haplotype Diversity and Linkage Disequilibrium at Human G6PD: Recent Origin of Alleles That Confer Malarial Resistance. *Science*, 293(5529), 455-462. <https://doi.org/10.1126/science.1061573>
- Wafer, R., Tandon, P., & Minchin, J. E. N. (2017). The Role of Peroxisome Proliferator-Activated Receptor Gamma (PPARG) in Adipogenesis: Applying Knowledge from the Fish Aquaculture Industry to Biomedical Research. *Frontiers in Endocrinology*, 8, 102. <https://doi.org/10.3389/fendo.2017.00102>
- Wakabayashi, K., Okamura, M., Tsutsumi, S., Nishikawa, N. S., Tanaka, T., Sakakibara, I., Kitakami, J., Ihara, S., Hashimoto, Y., Hamakubo, T., Kodama, T., Aburatani, H., & Sakai, J. (2009). The Peroxisome Proliferator-Activated Receptor  $\gamma$ /Retinoid X Receptor  $\alpha$  Heterodimer Targets the Histone Modification Enzyme PR-Set7/Setd8 Gene and Regulates Adipogenesis through a Positive Feedback Loop. *Molecular and Cellular Biology*, 29(13), 3544-3555. <https://doi.org/10.1128/MCB.01856-08>

Wang, J., Vasaiakar, S., Shi, Z., Greer, M., & Zhang, B. (2017). WebGestalt 2017: A more comprehensive, powerful, flexible and interactive gene set enrichment analysis toolkit. *Nucleic Acids Research*, *45*(W1), W130-W137.

<https://doi.org/10.1093/nar/gkx356>

Wang, L., Sinnott-Armstrong, N., Wagschal, A., Wark, A. R., Camporez, J.-P., Perry, R. J., Ji, F., Sohn, Y., Oh, J., Wu, S., Chery, J., Moud, B. N., Saadat, A., Dankel, S. N., Mellgren, G., Tallapragada, D. S. P., Strobel, S. M., Lee, M.-J., Tewhey, R., ... Näär, A. M. (2020). A MicroRNA Linking Human Positive Selection and Metabolic Disorders. *Cell*, *183*(3), 684-701.e14.

<https://doi.org/10.1016/j.cell.2020.09.017>

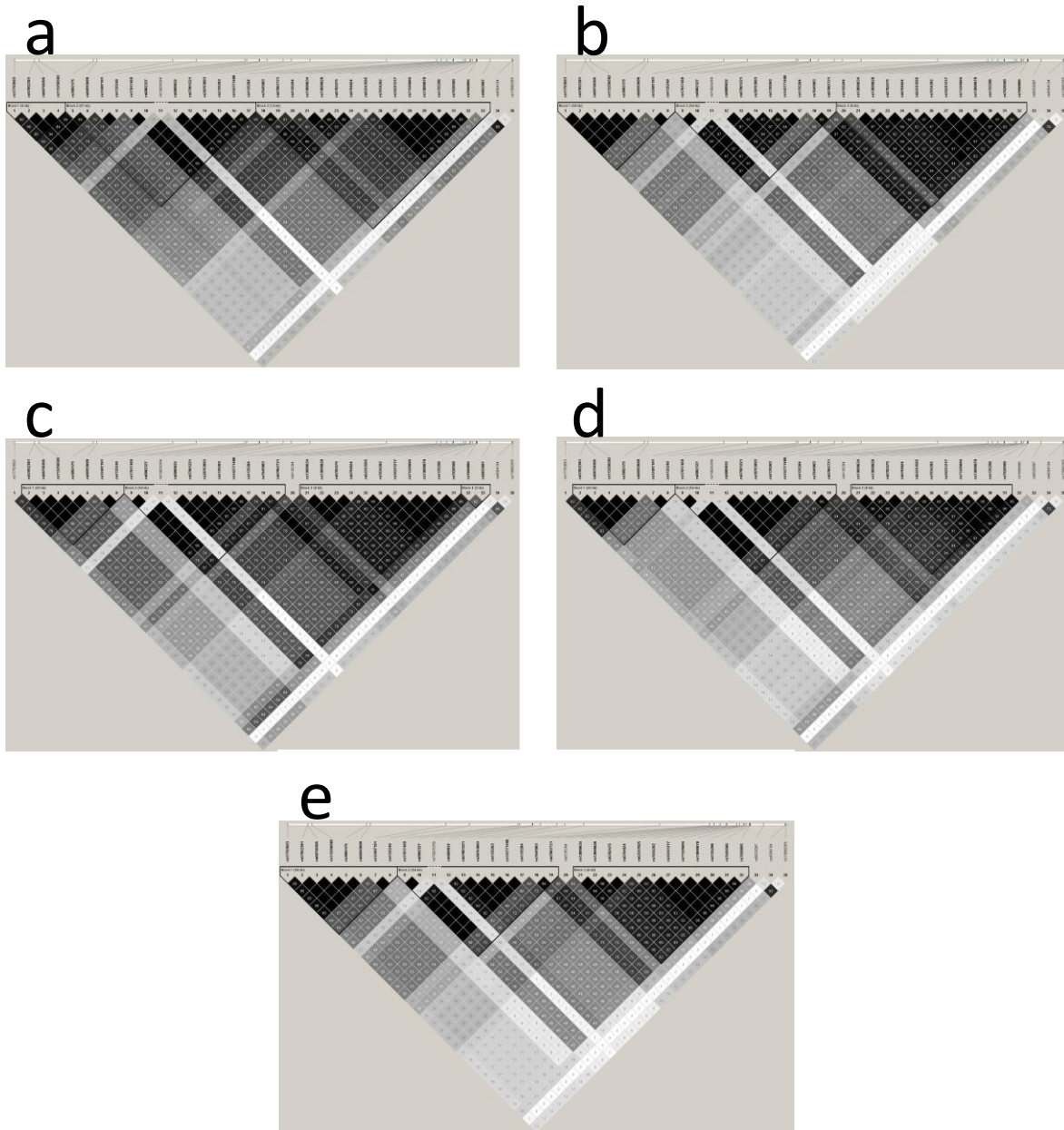
Yates, A. D., Achuthan, P., Akanni, W., Allen, J., Allen, J., Alvarez-Jarreta, J., Amode, M. R., Armean, I. M., Azov, A. G., Bennett, R., Bhai, J., Billis, K., Boddu, S., Marugán, J. C., Cummins, C., Davidson, C., Dodiya, K., Fatima, R., Gall, A., ... Flicek, P. (2019). Ensembl 2020. *Nucleic Acids Research*, gkz966. <https://doi.org/10.1093/nar/gkz966>

Zhou, Q., Zhao, L., & Guan, Y. (2016). Strong Selection at MHC in Mexicans since Admixture. *PLoS Genetics*, *12*(2), e1005847.

<https://doi.org/10.1371/journal.pgen.1005847>

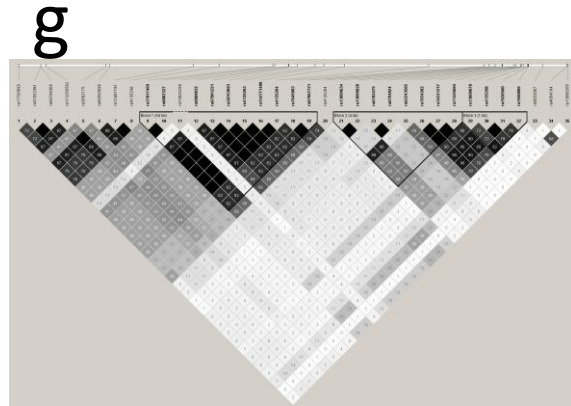
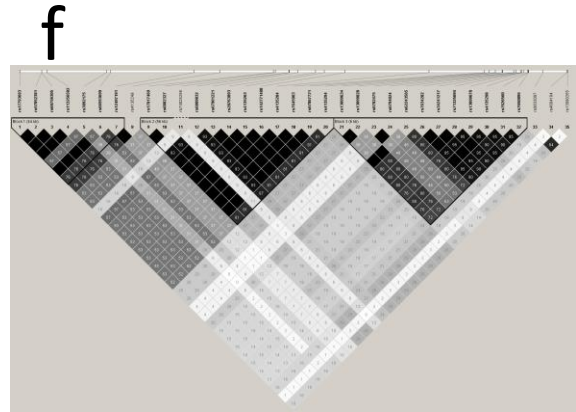
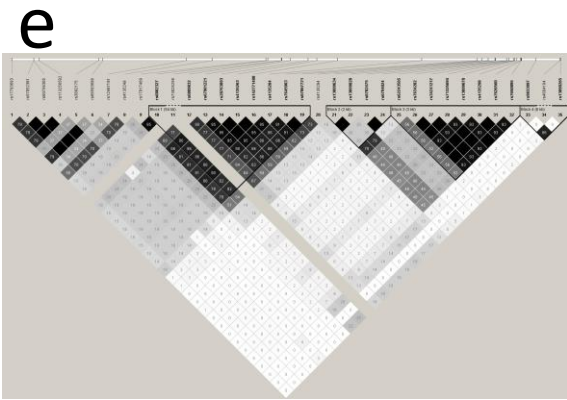
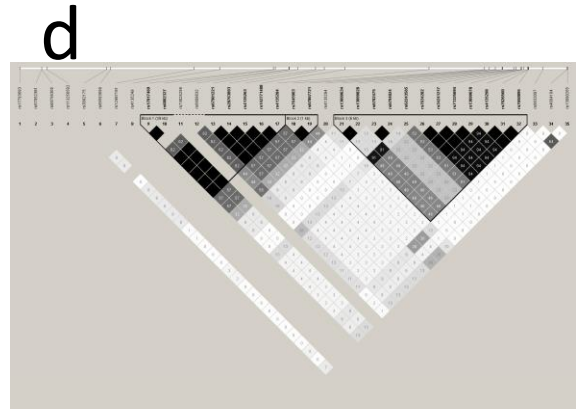
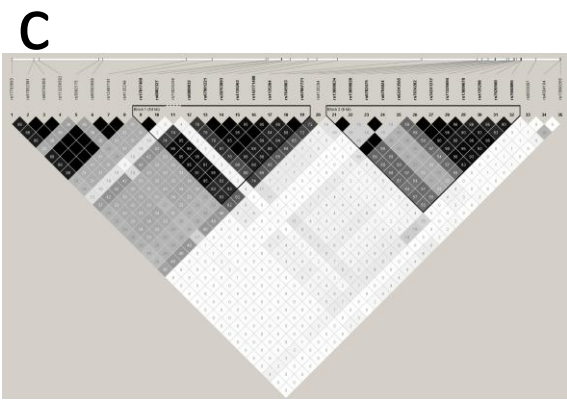
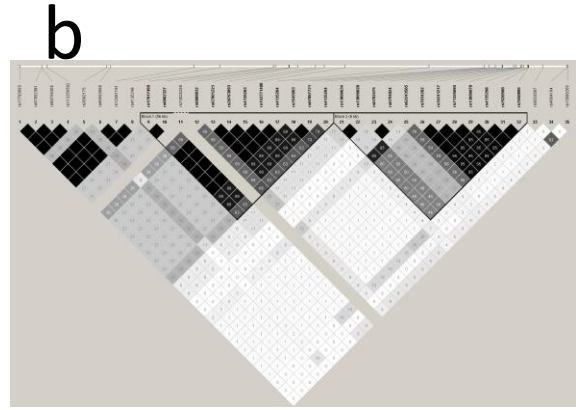
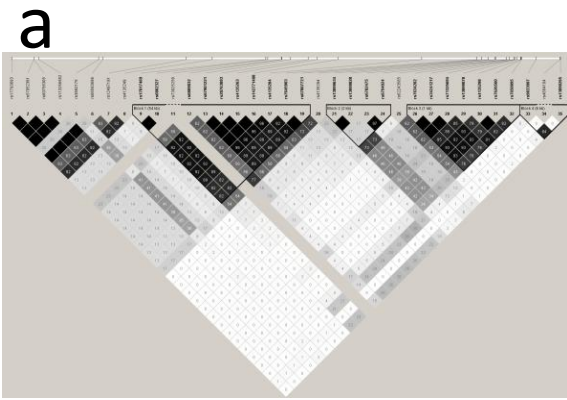
## X. Anexos

### 10.1 Gráficas de disequilibrio de ligamiento en el gen *PPARG* en las poblaciones de la fase 3 del proyecto de los 1000 genomas

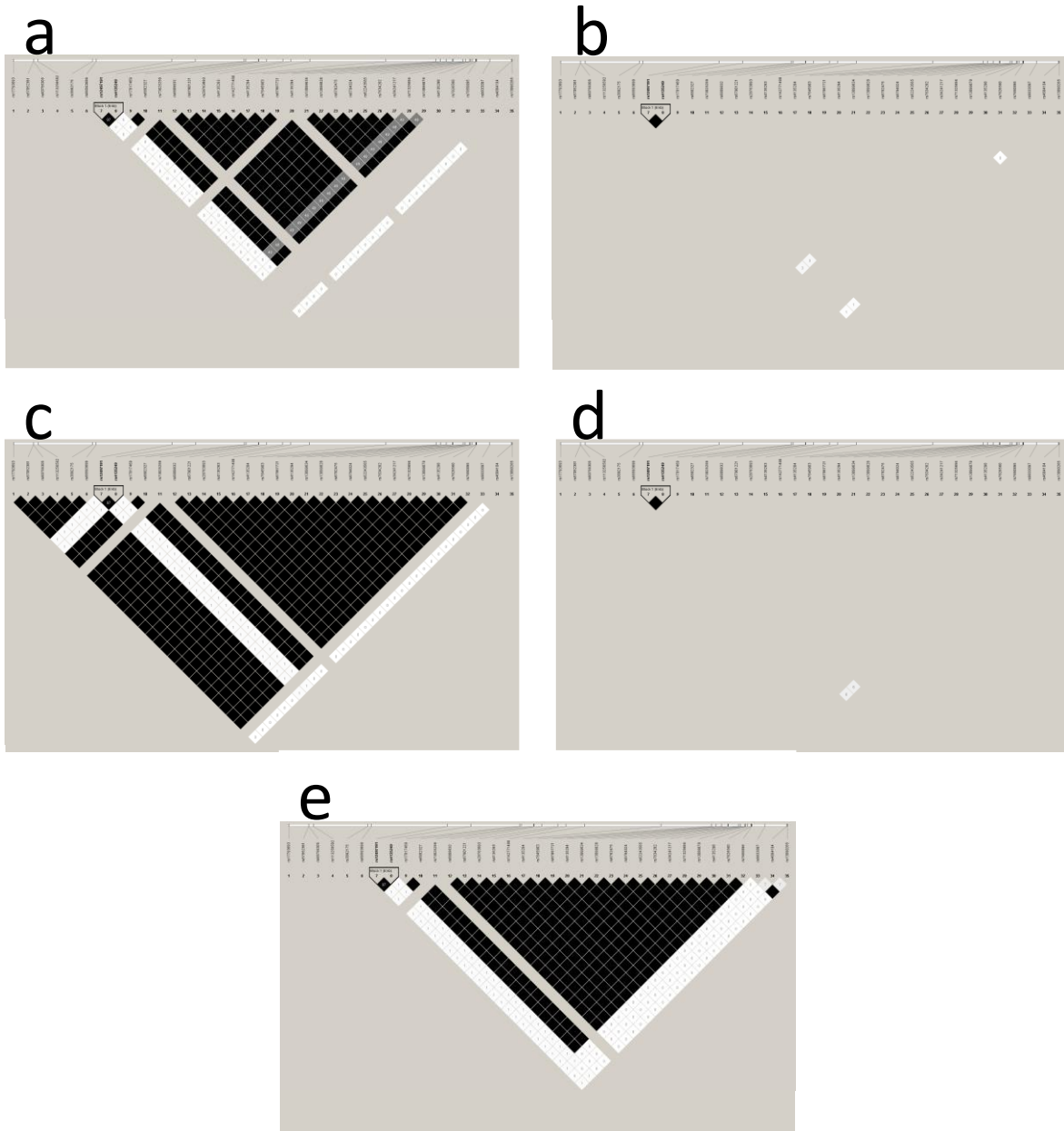


**Anexo 1 Poblaciones europeas del proyecto 1000 Genomas.** Disequilibrio de ligamiento de las variantes que componen el haplotipo bajo selección identificado en población indígena en poblaciones europeas derivadas del proyecto 1000 genomas. (a) Residentes de Utah con Ancestría del Norte y del Oeste de Europa (CEU), (b) Toscani, Italia (TSI), (c) Finlandeses, Finlandia (FIN), (d) británicos, Inglaterra y Escocia (GBR) y (e) Ibéricos, España (IBS)



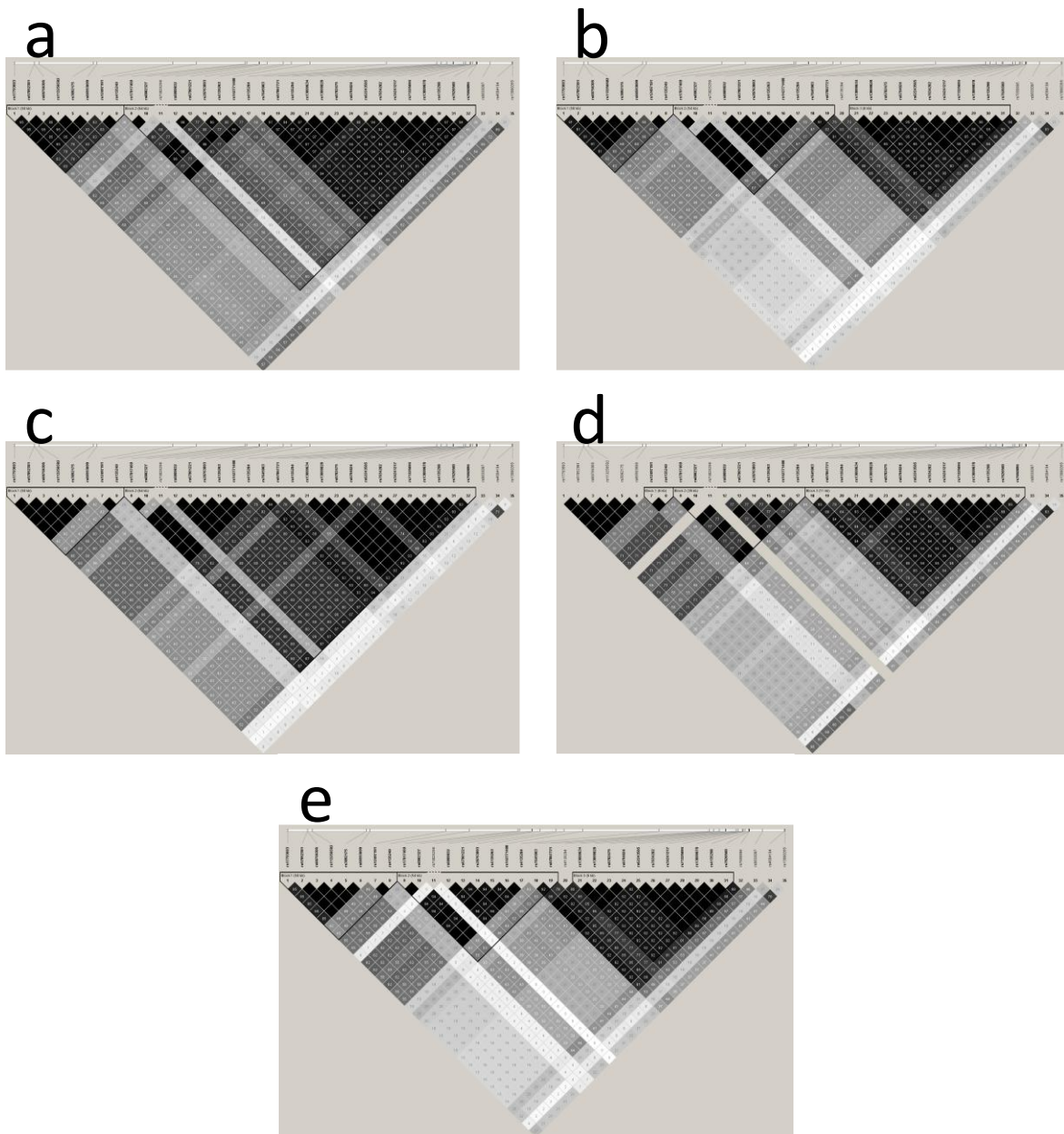


**Anexo 2 Poblaciones africanas del proyecto 1000 Genomas.** Desequilibrio de ligamiento de las variantes que componen el haplotipo bajo selección identificado en población indígena en poblaciones africanas derivadas del proyecto 1000 genomas. (a) Yoruba, Nigeria (YRI), (b) Luhya, Kenia (LWK), (c) Gambiaños en Divisiones del Oeste Gambia (GWD), (d) Mende, Sierra Leona (MSL), (e) Esan, Nigeria (ESN), (f) Americanos de Ancestría Africana, Sureste de Estados Unidos (ASW) y (g) Africanos Caribeños, Barbados (ACB)

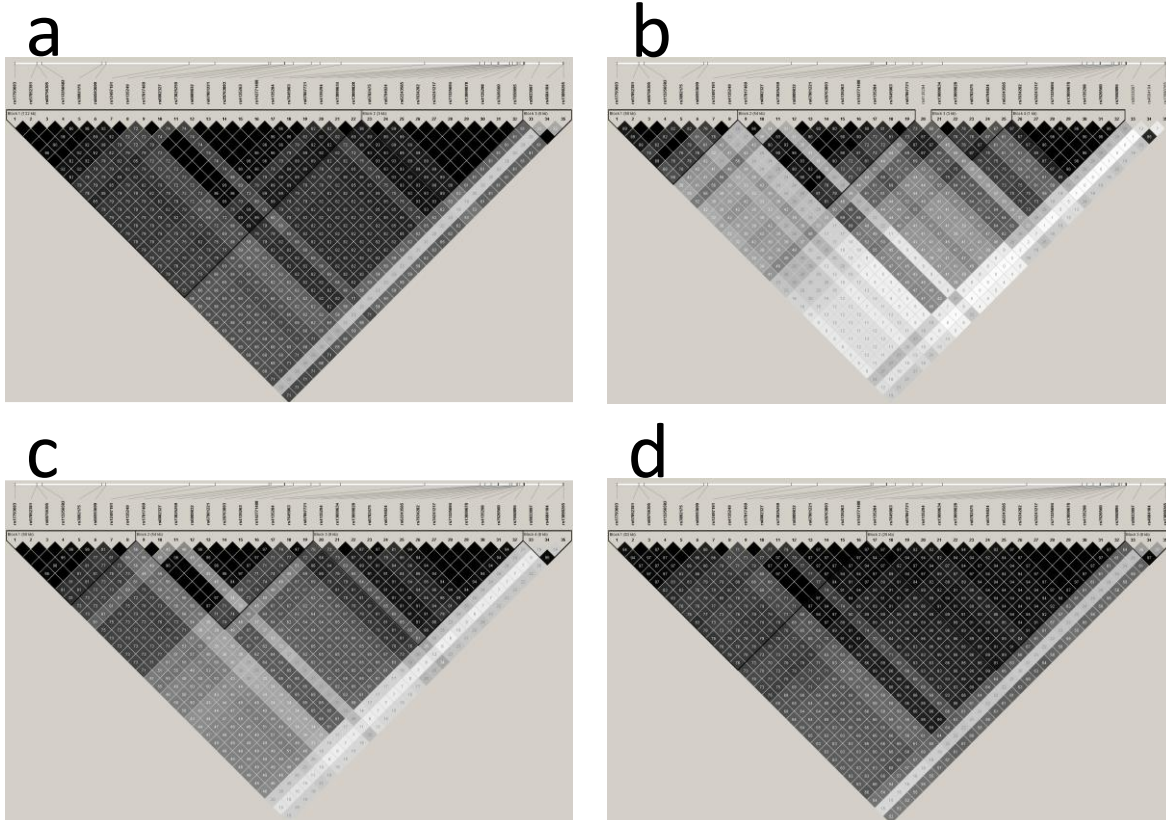


**Anexo 3 Poblaciones asiáticas del oriente del proyecto 1000 Genomas.** Desequilibrio de ligamiento de las variantes que componen el haplotipo bajo selección identificado en población indígena en poblaciones de Asia Oriental derivadas del proyecto 1000 genomas.

(a) Han, China (CHB), (b) Japoneses, Japón (JPT), (c) Han del Sur, China (CHS), (d) Dai, China (CDX) y (e) Kihn de la Ciudad Ho Chi Minh, Vietnam (KHV)



**Anexo 4 Poblaciones de asiáticas del sur del proyecto 1000 Genomas.** Desequilibrio de ligamiento de las variantes que componen el haplotipo bajo selección identificado en población indígena en poblaciones asiáticas del sur derivadas del proyecto 1000 genomas; (a) Indios Gujarati de Houston, Texas (GIH), (b) Punjabi de Lahore, Pakistán (PJL), (c) Bengalí, Bangladesh (BEB) (d) Sri Lankan Tamil del Reino Unido (STU) y (e) Indios Telugu del Reino Unido (ITU)



**Anexo 5 Poblaciones americanas del proyecto 1000 Genomas.** Desequilibrio de ligamiento de las variantes que componen el haplotipo bajo selección identificado en población indígena en poblaciones americanas derivadas del proyecto 1000 genomas. (a) Ancestría Mexicana de Los Ángeles, Estados Unidos (MXL) (b) Puertorriqueños, Puerto Rico (PUR), (c) Colombianos de Medellín, Colombia (CLM) y (d) Peruanos de Lima, Perú (PEL)