UNIVERSIDAD AUTÓNOMA DE BAJA CALIFORNIA
INSTITUTO DE INGENIERÍA

MAESTRÍA Y DOCTORADO EN CIENCIAS E INGENIERÍA

# MÉTODOS BIOINFORMÁTICO PARA LA IDENTIFICACIÓN DE VARIACIONES ESTRUCTURALES Y FACTORES DE RIESGO DE LA COMPOSICIÓN CORPORAL ASOCIADOS A LA HERNIA DE DISCO LUMBAR EN LA POBLACIÓN DE MÉXICO.

TESIS PARA RECIBIR EL GRADO DE:
***Doctorado en Ciencias***

PRESENTA:
***Adriana Griselda Mateos Valenzuela***

**Director:** Dr. Rafael Villa Angulo
**Co-Director:** Dr. Mirvana Elizabeth González Macías

Mexicali, B.C. Diciembre de 2024.

**AGRADECIMIENTOS**

*A Dios y a la vida por permitirme lograr un sueño más, en los tiempos perfectos.*

*A mi familia, por el amor, cariño y paciencia que me han tenido durante todo este proceso.*

*A mi sobrina Diana, sin ti esto no se logra. Mil gracias.*

*A mis amigas, Claudia por hacerme reír con las simpladas y por tus palabras de apoyo. Mirvana, por ver más allá y enseñarme que también lo puedo lograr*.

*A todos aquellos que estuvieron antes que yo y que forman parte de lo que soy.*

*A los Drs. Carlos y Rafael Villa por aceptarme nuevamente como estudiante de posgrado.*

*A CONAHCYT por el apoyo económico que me permitió continuar con mi preparación.*

*A UABC, porque es un orgullo ser estudiante y docente de la institución.*

**DEDICATORIA**

*A mis hijos, Fernando, José y Carlos, porque son y siempre serán mi mejor regalo. Engrandecen mi alma. Los amo.*

*A mí querida  y amada mamá, porque siempre me has dado lo mejor de ti, tu amor, tu tiempo y sabiduría.*

*A mí papá, por tus porras desde allá. Sé que estas muy orgulloso de mi. Esta es mi versión. Te amo.*

*A mí, porque me encontré y estoy feliz.*

**Sección** **Página**

**AGRADECIMIENTOS**

**DEDICATORIA**

**INDICE**

**LISTA DE FIGURAS**

**LISTA DE TABLAS**

**Contenido**

**CAPÍTULO I. INTRODUCCIÓN** ..... 9

**1.1 Introducción a la Genética.** ..... 9

**1.2 Genotipo y Fenotipo.** ..... 10

**1.3 Genética de poblaciones.** ..... 11

**1.3.1 Selección natural.** ..... 11

**1.3.2 Deriva Genética.** ..... 11

**1.3.3 Flujo genético.** ..... 12

**1.3.4 Mutación.** ..... 12

**1.3.5 Principio de Hardy-Weiberg.** ..... 13

**1.3.6 Desequilibrio por ligamiento (LD).** ..... 13

**1.4 Genética Humana.** ..... 14

**1.4.1 Variaciones Estructurales en el Genoma.** ..... 14

**1.4.2 Marcadores genéticos SNP.** ..... 14

**1.4.3 Variaciones estructurales basadas en LD de rango corto.** ..... 15

**1.4.4 Enfermedades hereditarias.** ..... 16

**1.4.5 Obesidad.** ..... 16

**1.4.6 Diabetes Mellitus tipo 2 (DM2).** ..... 17

**1.4.7 Hipertensión Arterial (HTA).** ..... 18

**1.4.8 Hernia de Disco Lumbar.** ..... 19

**1.4.9 Conformación de la población mexicana.** ..... 19

**1.5 Componentes de la Composición Corporal.** ..... 21

**1.5.1 Masa Grasa.** ..... 21

**1.5.2 Área de Grasa Visceral.** ..... 21

**1.5.2 Masa Magra.** ..... 22

**1.5.3 Agua corporal total.** ..... 22
**1.6 Métodos de Evaluación de la Composición Corporal.** ..... 22
**1.6.1 Antropometría.** ..... 22
**1.6.2 Bioimpedancia eléctrica (AIB).** ..... 23
**1.7 Factores de influyen en la composición corporal.** ..... 23
**1.7.1 Edad y género.** ..... 24
**1.7.2 Nivel de actividad física.** ..... 24
**1.7.3 Condición de salud.** ..... 25
**1.8 Microarreglos de genotipificación de salida masiva.** ..... 25
**1.8.1 GeneChip 500K array (Affymetrix)** ..... 26
**1.8.2. HumanHap550K array (Illumina)** ..... 26
**1.9 Estadística bioinformática** ..... 27
**1.9.1 Correlaciones de Pearson** ..... 28
**1.9.2 Prueba de t test para igualdad de medias** ..... 28
**1.9.3 Análisis de Componentes Principales (PCA)** ..... 29
**1.9.4 Regresión lineal** ..... 30
**1.9.5 Razón Odds** ..... 31
**1.10 Herramientas Bioinformáticas** ..... 32
**1.10.1 Lenguaje de Programación Python.** ..... 32
**1.10.2 Software Plink.** ..... 33
**1.10.3 Software R.** ..... 33
**1.10.4 Algoritmo para Detección de Variaciones Estructurales** ..... 33
**CAPÍTULO II. ANÁLISIS DE VARIACIONES ESTRUCTURALES BASADOS EN LD DE RANGO CORTO EN 10 POBLACIONES NATIVAS Y MESTIZAS DE MÉXICO**. ..... 36
**2.1Resumen** ..... 36
**2.2 Introducción** ..... 37
**2.3 Materiales y Métodos.** ..... 42
**2.3.1 Declaración de Ética para la obtención de datos.** ..... 42
**2.3.2 Descripción de los datos.** ..... 42
**2.3.3 Control de calidad de los datos.** ..... 43
**2.3.4 Desequilibrio por Ligamento.** ..... 43

**2.3.5 Variaciones Estructurales de Alta Densidad basadas en LD de Rango Corto.** ..... 44
**2.3.6 Corrección de Múltiples Pruebas** ..... 45
**2.3.7 Análisis de Componentes Principales.** ..... 45
**2.4 Resultados.** ..... 46
**2.4.1 Distribución de frecuencia de alelo menor.** ..... 46
**3.4.2 Análisis de LD.** ..... 47
**2.4.3 Variaciones Estructurales Basadas en LD de Rango Corto.** ..... 49
**2.5 Discusión** ..... 52
**CAPÍTULO III. FACTORES DE RIESGO Y ASOCIACIÓN DE COMPONENTES DE LA COMPOSICIÓN CORPORAL PARA LA HERNIA DE DISCO LUMBAR EN EL NOROESTE DE MÉXICO.** ..... 58
**3.1Resumen** ..... 58
**3.2 Introducción** ..... 59
**3.3 Materiales y Métodos** ..... 61
**3.3.1 Población de estudios.** ..... 61
**3.3.2 Recolección de Datos.** ..... 61
**3.3.4 Análisis de Componentes Principal (PCA)** ..... 63
**3.3.5 Análisis de regresión logística.** ..... 64
**3.3.6 Cálculo del factor de riesgo.** ..... 66
**3.4 Resultados.** ..... 67
**3.4.1 Análisis de detección de cúmulos.** ..... 67
**3.4.2 Análisis de asociación** ..... 68
**3.4.3 Análisis de factores de riesgo.** ..... 70
**3.5 Discusión.** ..... 71
**CAPITULO IV. CONCLUSIÓN** ..... 74
**REFERENCIAS** ..... 76

**LISTA DE FIGURAS** **Página**

Figura 1. Estructura de ADN……11
Figura 2. Mecanismos de selección natural……13
Figura 3. Microarreglo Affymetrix…….....28
Figura 4. Microarreglo Illumina…….....29
Figura 5. Gráfica de Distribución de Frecuencias Alélicas……..49
Figura 6. Gráfica de Desequilibrio de Ligamiento……50
Figura 7. Gráfica de PCA de VEs por cromosoma……53
Figura 8. Gráfica de PCA de Hernia…….....69

**LISTA DE TABLAS** **Página**

Tabla 1. Estadística de las Variación Estructurales Identificadas……..51

Tabla 2. Variables, Medias y su Desviación Estándar……..64

Tabla 3. Valores Significativos en la Regresión Lineal……71

Tabla 4. Razón Odds para HDL en Relación a Variables Significantes……72

# CAPÍTULO I. INTRODUCCIÓN

## 1.1 Introducción a la Genética.

La genética es la ciencia que permite explicar la herencia de generación en generación y la variabilidad humana, resultado del origen, migraciones ancestrales, evolución y la interacción con el medio ambiente del individuo o de la población.

La información que se requiere para el estudio de la genética se encuentra dentro del genoma humano, molécula formada por ADN (ácido desoxirribonucleico). Por su parte el ADN se encuentra empaquetado dentro de ciertas estructuras llamadas cromosomas. El ser humano está compuesto por 23 pares de cromosomas: 22 autosomales y el último par, el sexual. Estas moléculas y estructuras se encuentran dentro de cada una de nuestras células (National Human Genome Research Institute, n.d.)

El ADN es una molécula compuesta por cadenas de bases nitrogenadas, A (adenina), C (citosina), G (guanina) y T (tiamina), unidas a un azúcar y a un fosfato. Las cadenas de ADN dentro de una célula se estructuran en forma de doble hélice, y se generan con cierto orden específico. Estas se heredan de padres a hijos, en forma de genes; que son unidades de segmentos de estas cadenas, que transmiten información para la creación y desarrollo de diversas funciones de un organismo. El genoma humano tiene entre 20,000 y 25,000 genes. Aproximadamente 99% de la información genética es idéntica en todos los individuos de una especie, mientras que el 1% restante presenta diferencias y establece las características propias de cada individuo (Información Genética Básica - Cómo Entender La Genética -, n.d.).

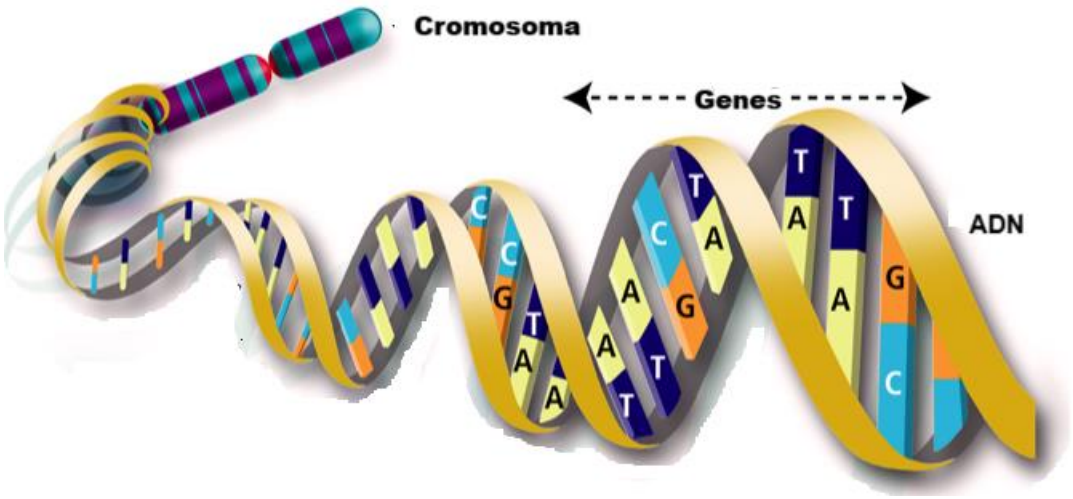


**Figura 1**. Estructuras del ADN.

**Nota**: Modificado de (U.S. Department of Energy Genomic Science program, n.d.).

### 1.2 Genotipo y Fenotipo.

El conjunto de genes heredados de padre a hijos conforma lo que se denomina genotipo. Cada individuo presenta dos versiones de cada gen, una heredada por la madre y otra por el padre. Cada gen tiene su configuración propia de ADN, a la cual llamamos Alelo; y cada alelo proporciona información de determinadas características que podrán verse reflejadas en el individuo, dependiendo de si son dominantes o si son recesivas. El genotipo nos proporcionará las instrucciones específicas de las características y el funcionamiento de nuestro organismo (National Human Genome Research Institute, n.d.).

Por otro lado, aquellas características que son observables se denominan fenotipo, como la altura, color de ojos, tono de piel, entre otros. Este no solo representa la información heredada, si no otros factores como los ambientales,

hábitos alimenticios, actividad física, estrés, entre otros. Estos factores podrán modificar el fenotipo de los individuos.

### 1.3 Genética de poblaciones.

Con el propósito de comprender el proceso evolutivo de las poblaciones, en este caso humanas, la genética de poblaciones estudia la variación y distribución de las frecuencias de los alelos entre todos sus individuos. Una población debe compartir el mismo medio en el que se desarrollan los individuos, y regirse principalmente por los siguientes mecanismos evolutivos: selección natural, deriva genética, flujo genético y mutación.

#### 1.3.1 Selección natural.

Es el mecanismo principal del proceso evolutivo, donde los individuos más adaptados son los que tienen mayor probabilidad de sobrevivir y de reproducirse, provocando cambios o variaciones sutiles en el fenotipo, y estas a su vez son heredadas registrando esta variación en las nuevas generaciones (Fornoni & Sosenski, 2009).

#### 1.3.2 Deriva Genética.

Es la fuerza evolutiva representado por la reproducción aleatoria de individuos, modificando las frecuencias de los alelos, produciendo cambios en el contenido genético de la población. Estos cambios en su variación alélica pueden reflejar un beneficio o prejuicio para la adaptación de las generaciones sucesivas.

### 1.3.3 Flujo genético.

Es el mecanismo de la evolución que está condicionado por el desplazamiento de genes de una población a otra. Este efecto se presenta en la migración, por ejemplo en la colonización de nuevas poblaciones, produciendo con ello el aumento de la variabilidad genética (Slatkin M, n.d.).

### 1.3.4 Mutación.

Durante el proceso de la división celular, el ADN genera una copia de él mismo. En algunas ocasiones esta copia presenta errores, estos errores se les denomina mutaciones, pueden ser ocasionadas por la exposición de sustancias con agentes mutágenos o por ciertas enfermedades ocasionadas por virus (National Human Genome Research Institute, n.d.).

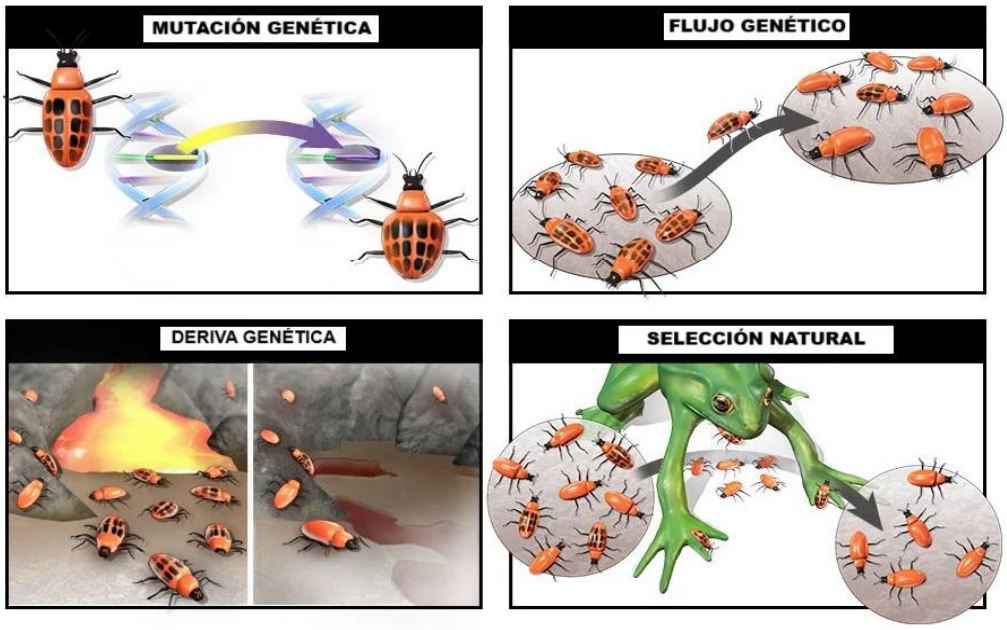


**Figura 2**. Mecanismo de Evolución.

**Nota**: Modificado de (VisibleBody, 2024).

### 1.3.5 Principio de Hardy-Weiberg.

El genotipo del ser humano se ha modificado a través de miles de millones años, transmitiendo en ellos variaciones en sus genes, frecuencia en alelos o en la estructura de los mismos, que contienen la información precisa para la adaptación y/o evolución, adecuándose al medio ambiente donde se desarrollan.

Estos cambios en las frecuencias de sus alelos son el resultado de 5 condiciones presentes en la mayoría de la población. Las condiciones son: mutaciones, apareamiento no aleatorio, población finita (deriva genética), selección natural y flujo de genético. Estas fuerzas producen el proceso evolución natural (Khan Academy, 2023).

Por lo contrario, cuando en una población no se presenta estas condiciones, y se sigue trasmitiendo la información genética sin cambios o variaciones, se dice que la población cumple con la ley o ecuación de equilibrio de Hardy-Weiberg, lo que equivale a una población sin evolución. Este hecho es llamado Principio de Hardy-Weinberg.

### 1.3.6 Desequilibrio por ligamiento (LD).

Es un indicador estadístico de la asociación no aleatoria de los alelos de dos o más posiciones base-par en alguna región específica de un cromosoma (Slatkin, 2016).

Este indicador permite conocer que combinaciones de genes determinan la expresión de ciertas características fenotípicas, mutaciones o algunas enfermedades propias de ciertas poblaciones o grupos, así como también el desarrollo, migración y evolución de la población.

## 1.4 Genética Humana.

### 1.4.1 Variaciones Estructurales en el Genoma.

Las variaciones estructurales (VEs) son el reordenamiento de segmentos de ADN, que involucran mínimo a 50 nucleótidos. Las VEs se reconocen como la forma más común de variación en bases par (bp) en el genoma humano. El efecto de las VEs está relacionado con numerosos mecanismos que alteran los genes codificantes de proteínas y la arquitectura cis-reguladora (Collins et al., 2020).

Los VEs se clasifican en reordenamientos “desequilibrados”, ganancias o pérdidas de ADN, como las Variaciones del Número de Copias (por sus siglas en inglés, CNV) y reordenamientos “equilibrados” los cuales no producen ninguna alteración (inversiones y translocaciones) (Chavez & Martinez, 2010).

Estos reordenamientos influyen en la configuración de la estructura y función del genoma. Impactando en la salud de los individuos. Existen VEs que están identificadas con ciertas patologías o condiciones como el cáncer, discapacidad cognitiva y la predisposición a la obesidad (Sudmant et al., 2015). Los VEs conllevan a la diferencia hereditaria de la expresión de los genes de la población humana (Abel et al., 2020).

### 1.4.2 Marcadores genéticos SNP.

El genoma humano contiene aproximadamente 3 mil millones de bases-par. Dentro de esta cadena de ADN se presentan ciertas variaciones en la secuencia, algunas de estas variaciones se pueden manifestar de manera individual alterando un solo nucleótido, a esto se le denomina polimorfismo de nucleótido simple (SNP, por sus siglas en ingles). Estas variaciones son las que se presentan más frecuentemente

a lo largo del genoma, aproximadamente cada 100 a 300 bases, en el genoma humano.

En la actualidad se han formado diferentes bases de datos con catálogos de SNPs que han sido identificados a lo largo del genoma, en estos catálogos se encuentran substituciones, microsatélites, repeticiones, inserciones y depleciones (Smigielski et al., 2000).

Algunas patologías o rasgos específicos pueden estar relacionados con marcadores genéticos, por lo que es necesario el estudio estas regiones específicas del ADN en las poblaciones (National Human Genome Research Institute, n.d.).

### 1.4.3 Variaciones estructurales basadas en LD de rango corto.

Las variaciones estructurales (VEs) pueden ser identificadas y estudiadas por el desequilibrio por ligamiento (LD, por sus siglas en inglés) de rango corto. En este trabajo de tesis en particular, usamos la definición hecha por Salomón-Torres et al., 2014; quien las define e identifica de la siguiente manera: 1) tomar el conjunto de SNPs y ordenarlo por posición de menor a mayor; 2) identificar y usar sólo SNPs que tengan al menos 15 SNPs vecinos en un rango de 100 kb; 3) calcular el desequilibrio por ligamento ($r^2$) de rango corto (≤100 Kb) de cada SNP con sus vecinos; 4) ordenar los pares de SNPs por su distancias y usando rangos de 5 kb obtener la media de $r^2$ de cada rango; 5) Inspeccionar todos los SNPs buscando segmentos de al menos 1kb de tamaño, que contengan al menos 3 SNPs adyacentes cuya $r^2$ con todos su vecinos, estén todas por arriba o todas por debajo de la media esperada de $r^2$ de sus respectivo rango, y sus *valores p* de la prueba-*t*,

para igualdad de medias, sean significativas después de una corrección de Benjamini-Hochberg (Salomón-Torres et al., 2014).

### 1.4.4 Enfermedades hereditarias.

La herencia juega un rol fundamental en la aparición de enfermedades que están asociadas con alteraciones los genes. En el contexto de las enfermedades hereditarias, las alteraciones genéticas se transmiten de generación en generación, propagando la enfermedad o su predisposición. En algunos casos son debido a la mutación de un solo gen (monogénicas), como el caso de la acondroplastia, enfermedad que está relacionado con la mutación del gen FGFR3 (Castaño & Bilbao, 1997).

Cuando una enfermedad es causada por la alteración de más de un gen, se les denomina poligénicas. Mientras que a las enfermedades que se producen por la interacción de factores ambientales y un conjunto de factores genéticos se les denomina enfermedades multifactoriales. Dentro de estas últimas podemos encontrar los padecimientos comunes, como la diabetes mellitus, ciertos tipos de cáncer, la obesidad, hipertensión y las hernias de disco lumbar (Genetic Alliance, 2008).

### 1.4.5 Obesidad.

La Organización Mundial de la Salud la define a la obesidad como, una enfermedad crónica que se caracteriza por la acumulación anormal, o excesiva, de tejido adiposo en relación con el peso, provocando varios procesos pro-inflamatorios que son perjudiciales para la salud (Organización Mundial de la Salud, 2021).

La etiología de esta enfermedad es multifactorial, generalmente resultado de la interacción de determinados genes susceptibles a la obesidad con el desequilibrio de energía debido a un estilo de vida sedentario, la ingesta de alimentos con alto contenido calórico, o ambos (González Jiménez, 2011).

En la actualidad se han descrito más de 1100 SNPs relacionados con la obesidad, algunos de estos SNPs afectan genes están implicados en la regulación de distintos procesos como: el gasto energético, el crecimiento, diferenciación e inflamación del tejido adiposo y el metabolismo de la insulina (González Jiménez, 2011). Sin embargo, solamente 12 SNPs han sido identificados como ambiento-genes, que son aquellos que por la interacción con el medio ambiente pueden exacerbar o disminuir el efecto, como: el gen FTO (Busebee et al., 2023).

El diagnóstico y grado de obesidad se evalúa en la práctica clínica mediante el Índice de Masa Corporal (IMC) que se correlaciona con la adiposidad. El punto de corte para definir la obesidad es el IMC ≥30.

### 1.4.6 Diabetes Mellitus tipo 2 (DM2).

La DM2 es una enfermedad metabólica compleja, de etiología multifactorial, caracterizada por la resistencia a la insulina y deficiencia relativa en la acción y secreción de dicha hormona, provocando un aumento en los niveles de glucosa en sangre. La DM2 es la forma más frecuente de diabetes, presentándose principalmente en población adulta.

Dentro de los factores de riesgo para el desarrollo de esta enfermedad son el sobrepeso, la obesidad, el sedentarismo, la dieta, hipertensión, edad y etnia. Las

etnias que se consideran con mayor predisposición para el desarrollo de esta condición son: estadounidenses de raza negra y oriental, latinos, indígenas estadounidenses e isleños del Pacífico (Organización Panamericana de la Salud, 2023).

Estudios genéticos han establecido a la DM2 como una enfermedad poligénica, encontrando en la población mexicana la asociación de la DM2 con genes como: el CAPN10, APOE, BGLAP, SCARB1, INS, IRS1, ELMO1 TCF7L2, entre otros (Ramírez-García et al., 2015).

### 1.4.7 Hipertensión Arterial (HTA).

Enfermedad multifactorial que se caracteriza por la elevación de la presión de la sangre en los vasos sanguíneos (≥140/90mmHg). Este incremento del rango normal de la HTA impacta la salud cardiovascular, convirtiendo esta enfermedad en una de las principales causas de morbi-mortalidad en nuestro país (Ramírez-Bello et al., 2011).

Dentro de los factores desencadenantes no modificables se encuentran los genéticos, la raza/etnia, edad y el género, mientras que dentro de los modificables están el sobrepeso y la obesidad, la falta de actividad física, alto consumo de sal, tabaco y alcohol, entre otros (Organización Mundial de la Salud, 2023).

La HTA contiene un componente genético. Algunos de los genes que están asociados a esta enfermedad son AGT, CALM2, NCX1, ADRA1, ADRB1, PLN, ESR, ANPEP, MTHF y MED1 (Patel et al., 2017; Pérez-Razo et al., 2015; Ramírez-Bello et al., 2011).

### 1.4.8 Hernia de Disco Lumbar.

La hernia de disco lumbar es una afección causada por la extrusión del interior del disco intervertebral (núcleo pulposo). Comprimiendo de manera mecánica, o debido a la inflamación, la irritación del nervio y ocasiona dolor de la espalda baja, que se irradia sobre el dermatoma de la pierna del nervio afectado. Su relevancia se debe a que es la segunda causa de incapacidad laboral y la evolución no satisfactoria perjudica notablemente en la calidad de vida del paciente (Ricci et al., 2006).

La etiología de la hernia de disco es multifactorial, consecuencia de la interacción de factores como la edad, carga mecánica, algún traumatismo, obesidad, el género y procesos que alteren el metabolismo como la DM2.

Por otra parte, estudios recientes sugieren que ciertos genes contribuyen a la susceptibilidad de las enfermedades del disco lumbar, como la herniación de éste. Se ha estableciendo como factores de riesgo distintos polimorfismos como el BDNF, IL6, IL10, CILP, TNFA, analizadas en poblaciones caucásicas, asiáticas y europeas (Eskola et al., 2012; Le Maitre et al., 2007; Liu et al., 2022; Zhu et al., 2018).

### 1.4.9 Conformación de la población mexicana.

Existen dos teorías en relación a los grupos humanos que poblaron América: el primero es la teoría del poblamiento temprano, que se remonta a 30,000 a 25,000 años en el pasado: y la segunda es teoría de poblamiento tardío, que sucedió hace 12,000 a 14,000 años. Se reconoce que los pobladores ancestrales provenientes de Siberia, fueron resultado de una mezcla de antiguos norsiberianos y asiáticos. A su vez, existen estudios de antropología biológica que sugieren la posibilidad de la relación con poblaciones Australo-Melanesios. Dicha población atravesó por la

Beringia, puente terrestre que existió entre Asia y América durante la última glaciación. Algunos de estos grupos ancestrales se establecieron en lo que ahora es Alaska y Yukón, mientras otros se dispersaron por todo el continente siguiendo una ruta terrestre o marítima (Moreno-Mayar et al., 2018; Raghavan et al., 2015; Reich et al., 2012).

Estos grupos humanos fueron los fundadores de los grupos nativos americanos (NA), que poco a poco se adecuaron a las distintas circunstancias locales para suministrase de alimento. Algunos mediante la recolección de plantas, y caza de pequeñas especies, otros con la caza de grandes especies y algunos más pescadores. De esta manera comenzaron a constituirse como sociedades agrícolas, sedentarias con organización tribal (Villegas et al., 1994).

En nuestro territorio, los primeros asentamientos de poblaciones nativas datan aproximadamente 2000 a.C. Estas se establecieron en regiones geográficas específicas en donde, aprovechando las características fisiográficas que influyen en las condiciones climáticas, el tipo de suelo, la vegetación, y la fertilidad de la tierra, brindaron la oportunidad para que los grupos nativos realizaran la recolección de su alimento; seguido de la siembra de cultivos como el maíz, calabaza, chile, frijol y cacao, adaptando su forma de vida y sobreviviendo a las distintas condiciones presentes en su entorno. Esta evolución produjo la regionalización y las constitución de ciudades (INEGI, 2022; Navarrete Linares, 2008; Zolla & Márquez, 2004).

A la información genética contenida por los grupos nativos; con la conquista de América por los españoles, se sumó la de las poblaciones extranjeras, principalmente europeas y en una mínima porción africana. Igualmente han

contribuido otros procesos históricos, epidemiológicos y de migración. Es por ello, que la “raza” mexicana actual, se caracteriza por una alta diversidad genética, representada por el grupo de los nativos, mestizos y afrodescendientes (Guardado-estrada et al., 2008).

## 1.5 Componentes de la Composición Corporal.

La dieta, el ejercicio físico, enfermedades y el crecimiento son factores que alteran el organismo, reflejándose en la composición corporal de los individuos. El análisis de los componentes corporales, permite brindar información sobre el estado de salud de los individuos mediante la cuantificación de la masa grasa, la masa músculo esquelética, el porcentaje de grasa, y el área de grasa visceral. Mostrando con ello una mayor precisión en relación a la salud en comparación con el IMC o el peso (Kyle et al., 2003).

### 1.5.1 Masa Grasa.

La masa grasa es el tejido adiposo acumulado debajo de la piel (tejido graso componente esencial de reserva energética y aislante nervioso (González Jiménez, 2013).

### 1.5.2 Área de Grasa Visceral.

El área de grasa visceral es el tejido adiposo se encuentra de manera subcutánea o en las cavidades corporales que envuelven a los órganos o vísceras. La acumulación de grasa en las vísceras esta relaciona con alteraciones metabólicas y hemodinámicas, además del incremento de niveles de citoquinas inflamatorias, convirtiendo este componente en un detonador de efectos perjudiciales a nivel cardio-metabólico (Almeida et al., 2018).

### 1.5.2 Masa Magra.

La masa magra es el peso total del cuerpo menos la masa grasa identificada. Esto equivale a todo aquel tejido diferente a la masa grasa como lo son los órganos, la piel, huesos la masa muscular y el agua corporal (InBody Co., 1996).

### 1.5.3 Agua corporal total.

El agua corporal total es el agua que se encuentra dentro del cuerpo humano representa entre la mitad y dos tercios del peso de una persona, dividiéndose en agua intracelular y extracelular. La intracelular refiere aquella que indica la cantidad de agua que existe dentro de las membranas celulares, mientras que la extracelular constituye la cantidad de agua que está dentro del fluido intersticial y la sangre. El tejido adiposo contiene un menor porcentaje de agua a comparación del tejido magro. Los porcentajes de agua en el individuo difieren según la edad y género (InBody Co., 1996).

## 1.6 Métodos de Evaluación de la Composición Corporal.

La evaluación de la composición corporal reside en la importancia de conocer los efectos que estos tienen en la salud, principalmente causa de la relación entre la cantidad de masa grasa y masa muscular. Dentro de los métodos de análisis se encuentran los indirectos (resonancia magnética, tomografía computarizada, la absorción de rayos X de energía dual y la plestismografía) y los doblemente indirectos (antropometría y la bioimpedancia) (Paredes, 2022).

### 1.6.1 Antropometría.

La antropometría es el estudio cuantitativo de las características físicas del ser humano. Por ejemplo, el peso es el valor medido como resultado de los diferentes

tejidos que conforman el cuerpo, mientras la talla es el valor medido de la altura con la que cuenta el individuo. Los valores de peso y talla son medidas antropométricas corporales de gran utilidad y de fácil obtención. Son utilizadas de manera rutinaria para evaluar el crecimiento y la epidemiologia.

El IMC, el cual es determinado por la relación entre el peso y la talla, es utilizado por la Organización Mundial de la Salud (OMS) para clasificar el estado de adiposidad y nutricional de sujetos. Siendo evaluados aquellos con un IMC entre 18.5 y 24.9 como normopesos, de 25 a 29.9 con sobre peso y obesidad tipo I, y aquellos con valores por igual o mayor a 30 con obesidad tipo II y tipo III a personas con valores de 35 a 39.9 (González Jiménez, 2013).

### 1.6.2 Bioimpedancia eléctrica (AIB).

La medición de la AIB es una herramienta que utiliza la metodología tetrapolar (en este caso) que permite analizar la composición corporal de manera no invasiva, económica y fácil operación. Utilizando una corriente alterna imperceptible, realiza la medición de la masa grasa y de la masa libre de grasa. Basándose en los valores obtenidos de impedancia, resistencia y reactancia corporal resultado de la conductividad del agua en los diferentes compartimientos del cuerpo. Seguido de ecuaciones predictivas que permiten el cálculo de la masa muscular (InBody Co., 1996; Paredes, 2022).

## 1.7 Factores de influyen en la composición corporal.

Existen diferentes factores que alteran la composición corporal, como lo es la edad, el género, factores genéticos, nivel de actividad física y la condición de salud.

### 1.7.1 Edad y género.

Estudios mencionan que las personas adultas, tanto en hombres como mujeres con edades de entre 50 y los 60 años muestran un incremento en el promedio de IMC. Sin importar el grupo étnico al que pertenecen.

Estos incrementos se ven de igual manera en los componentes de la masa grasa total y en porcentaje de grasa. Por lo contrario, estos valores se ven disminuidos en personas mayores, entre los 70 y 80 años.

Por otro lado, se ve una disminución de la talla a partir de los 40 años, reflejándose esto en un incremento del IMC. Esta modificación se aprecia más drásticamente en las mujeres postmenopáusicas (Anders Aandstad, Kristian Holthergerget, Rune Hageberg, Ingar Holme, 2014).

En relación al tejido adiposo, los hombres presentan una mayor cantidad de grasa visceral que las mujeres en cualquier etapa de vida, incrementándose este nivel con la edad. Mientras que la masa magra también se ve afectada, sobre todo la musculo esquelética, apreciándose una disminución de este componente con el paso de los años.

### 1.7.2 Nivel de actividad física.

La actividad física desencadena una serie de procesos benéficos para la salud. Derivado de los diferentes efectos que tienen a nivel hormonal y metabólico en el cuerpo humano.

La actividad física está asociada con una menor cantidad de masa grasa total y visceral  a pesar de que pueda existir una predisposición a la obesidad (Woon et al., 2020).

### 1.7.3 Condición de salud.

Algunas de las condiciones de salud que pueden afectar la composición corporal son principalmente las relacionadas con la nutrición, como es la desnutrición y la obesidad. Otros padecimientos importantes que afectan los componentes son la osteoporosis, y la sarcopenia. En el caso de la osteoporosis se ve afectado la cantidad de minerales, mientras en la sarcopenia la masa magra se verá disminuida a la vez que en algunos casos puede incrementarse el nivel de masa grasa.

Ciertas condiciones que presentan los pacientes con cáncer como la caquexia, y el linfedema, también afectan la composición corporal. En el caso de la caquexia se ve reflejado en una disminución de peso y de masa muscular esquelética, mientras que el linfedema provoca un aumento en algunos de los miembros, brazos o piernas, debido a la acumulación de linfa en estos segmentos, incrementando los niveles de agua extracelular en la zona específica (Holmes & Racette, 2021).

## 1.8 Microarreglos de genotipificación de salida masiva.

Es una herramienta tecnológica, conocida como SNPchip o SNP array que permite el análisis simultáneo de miles o millones de polimorfismos. Con esta tecnología, es posible realizar estudios de asociación para identificar rasgos, características o

enfermedades que estén relacionados con marcadores genéticos (National Human Genome Research Institute, n.d.).

### 1.8.1 GeneChip 500K array (Affymetrix)

Los GeneChip son microarreglos de oligonucleótidos de alta densidad. Estos miden la expresión a nivel genómico-transcriptómico mediante la técnica de hibridación específica de ARN-ADN o ADN-ADN de múltiples genes de manera simultánea, utilizando en la base del arreglo, sondas con secuencias de transcriptoma, para detectar la región complementaria estudiada (Jiang et al., 2007; Universidad de Santiago de Compostela, 2024).

**Figura 3.** Microarreglo de Affymetrix.
**Nota**: Recuperado de (Jeff Smith, n.d.).

### 1.8.2. HumanHap550K array (Illumina)

Los SNPchips de la empresa Illumina son una herramienta tecnológica avanzada que se utiliza en la investigación para el análisis de datos genómicos. Esta tecnología, llamada Beadarray, se caracteriza por utilizar micro esferas de sílice, las

cuales analizan de manera simultánea cientos de miles o millones de SNPs de un solo individuo. Al pasar cada fragmento de ADN de la muestra, las sondas se van uniendo a la secuencia complementaria, fijándose en la base anterior del marcador de interés. Al ser excitadas por medio de un láser, emiten una determinada intensidad de señal otorgando la información sobre la proporción alélica de ese locus(illumina, n.d.).

**Figura 4.** Microarreglo de Illumina
**Nota:** Recuperado de (illumina, 2009).

### 1.9 Estadística bioinformática

Termino que involucra de manera multidisciplinaria distintas áreas de la ciencia como, la biología computacional, informática, matemáticas y estadística, para el almacenamiento, recuperación, organización y análisis de datos biológicos. Los bioinformáticos, mediante el uso de distintas técnicas, métodos y aplicación de herramientas tecnológicas, identifican estadísticamente los cambios significativos de los procesos biológicos(Jeffrey S. Morris and Veerabhadran Baladandayuthapani, 2017).

### 1.9.1 Correlaciones de Pearson

La Correlación de Pearson es una medida de dependencia lineal entre dos variables aleatorias cuantitativas. Generalmente es estimado el coeficiente de correlación de Pearson para medir el grado de relación entre dos variables, siempre y cuando ambas sean cuantitativas y continuas.

Cuando este coeficiente es aplicado a muestras, suele denotarse como $r_{xy}$, y es comúnmente llamado como *Coeficiente muestral* o *Coeficiente de correlación muestral de Pearson.* La forma de calcularlo es mediante la siguiente formula:

$$r_{xy} = \frac{\sum_{i=1}^{n}(x_i - x_{media})(y_i - y_{media})}{\sqrt{\sum_{i=1}^{n}(x_i - x_{media})^2}\sqrt{\sum_{i=1}^{n}(y_i - y_{media})^2}}$$

donde:

*n* es el tamaño de la muestra.

$x_i$, $y_i$ Son puntos muestrales individuales indexados con *i.*

$x_{media}$, $y_{media}$ son la media de cada muestra $x_i$ y $y_i$ respectivamente.

### 1.9.2 Prueba de t test para igualdad de medias

Prueba paramétrica, la cual examina si la media de una población es estadísticamente diferente de un valor conocido o hipotético.

La fórmula estadística de una sola muestra para la prueba *t* se expresa de la siguiente manera:

$$t = \frac{\overline{x} - \mu_0}{s_{\overline{x}}}$$

En donde:

$\mu_0$ = La constante para la media de la población

$\bar{X}$ = La media de la muestra

*n* = El tamaño de la muestra

*s* = La desviación estándar de la muestra.

$S_{\bar{X}}$ = El error estimado de la media

El valor *t* calculado es comparado con el valor crítico de la distribución *t* con grados de libertad *df*=n-1. Con un nivel de significancia de 0.05, Si el valor *t* calculado es mayor que el valor crítico, entonces se rechaza la hipótesis nula y se concluye que las medias no son iguales, de lo contrario se acepta la igualdad de las medias (Mishra et al., 2019).

### 1.9.3 Análisis de Componentes Principales (PCA)

Técnica estadística multivariante de simplificación, la cual hace referencia a la transformación óptima de un conjunto de datos correlacionados dados, a un conjunto sintético no correlacionados denominados componentes principales. Mediante esta técnica se elimina la información repetida y se ordena las nuevas variables de acuerdo con la dispersión de la información total con la que se cuenta (Jolliffe, 2002).

El procedimiento para obtener los componentes principales puede resumirse de la siguiente manera: dado un vector $\mathbf{X^T}$ de *n* dimensiones, $X = [x_1, x_2 \ldots, x_n]^T$, del cual sus vectores de medias, **M,** y covarianzas, **C,** son descritos por: $M=E(X)= [m_1, m_2, \ldots, m_n]^T$ y $C=E[(X-M)(X-M)^T]$; calcular los eigenvalores $\lambda_1, \lambda_2, \ldots, \lambda_n$, y los

eigenvectores $P_1, P_2, \ldots, P_n$; y ordenarlos de acuerdo a su magnitud $\lambda_1 \geq \lambda_2 \geq \ldots \geq \lambda_n$. Seleccionar $d$ eigenvectores para representar las $n$ variables, $d<n$. Entonces $P_1, P_2, \ldots, P_d$ son llamados componentes principales.

### 1.9.4 Regresión lineal

Los Modelos de Regresión Múltiple son métodos matemáticos para modelar la relación estocástica cuantitativa entre una variable de interés y un conjunto de variables explicativas. De forma general estos modelos pueden ser expresados como sigue:

$$Y_i = \beta_0 + \beta_1 X_{1i} + \beta_2 X_{2i} + \ldots + \beta_p X_{pi} + \varepsilon_i$$

donde:

$Y_i$ : variable de interés, dependiente o regresando,

$X_{1i}, X_{2i}, \ldots, X_{pi}$ : variables explicativas, independientes o regresores,

$\beta_0$ : intersección o termino constante,

$\beta_1, \beta_2, \ldots, \beta_p$ : parámetros, miden la influencia que las variables explicativas tienen sobre el regresando,

$p$ : número de parámetros independientes a tomar en cuenta,

$\varepsilon$ : error de observación debido a variables no controladas,

$i$ : $1, 2, \ldots, n$ : número de observaciones de las variables.

Con estos modelos es posible estudiar relaciones lineales entre múltiples variables y el efecto que estas tienen sobre la variable dependiente. Los $\beta_i$ se estiman siguiendo el criterio de mínimos cuadrados:

$$\min_{\substack{\beta \in n \\ i=1,\ldots,n}} \sum_{i=1}^{n} \left( Y_i - \beta_0 - \beta_1 X_{1i} - \beta_2 X_{2i} - \ldots - \beta_p X_{pi} \right)^2$$

Los estimadores de mínimos cuadrados son obtenidos a partir de la ecuación:

$$\widehat{\beta} = \left(X^T X\right)^{-1} X^{-T} Y$$

### 1.9.5 Razón Odds

La razón Odds (OR) es una medida de asociación entre una variable causante y una variable respuesta. El OR representa la probabilidad de que un evento ocurra dado una exposición particular, en comparación con la probabilidad de que el resultado se produzca en ausencia de esa exposición (Szumilas, 2010).

El OR también se puede utilizar para determinar si una exposición en particular es un factor de riesgo para un resultado y para comparar la magnitud de varios factores de riesgo para ese resultado.

En el contexto de regresión logística, el coeficiente de regresión (CR) es un estimador del incremento logarítmico de la razón de probabilidad de la variable respuesta (incremento del log odd) cuando incrementa en el 1 el valor de la variable causante, por lo que: la OR es el exponencial del coeficiente de regresión ($e^{CR}$), y corresponde a la razón de probabilidad asociada al incremento de una unidad en la variable causante. Los criterios para definir los factores de riesgo basados en la OR son los mencionados por Szumila, 2010. Donde establece lo siguiente:

Si OR = 1 La variable causante no afecta el valor de la variable respuesta; no existe asociación entre las variables,

Si OR > 1 La variable causante está asociada con probabilidades altas de la variable respuesta; existe asociación entre las variables,

Si OR < 1 La variable causante está asociada con probabilidades bajas de la variable respuesta; existe asociación protectora entre las variables.

Aunado a esto, el intervalo de confidencia (CI) se utiliza para estimar la precisión del OR. Un valor de CI alto indica un bajo nivel de precisión de la OR, mientras que un valor bajo de CI indica una mayor precisión del OR. En la práctica si el CI no contiene el valor 1, que es el valor nulo entonces decimos que la OR es significante estadísticamente (este criterio es usado como una aproximación).

## 1.10 Herramientas Bioinformáticas

Son programas matemáticos y computacionales que permiten realizar de manera eficiente el manejo y análisis de datos biológicos. La biología del siglo XXI exige la manipulación y procesamiento de grandes cantidades de datos por lo que es necesario el empleo de formatos y algoritmos computacionales, así como, de bases de datos para lograr una correcta interpretación de los mismos (Bayat, 2002).

### 1.10.1 Lenguaje de Programación Python.

Python(Python Software Foundation., 2001) es un lenguaje de programación libre y gratuita, creado por Guido van Rossum en 1990. Se caracteriza por ser un lenguaje de alto nivel, lo que permite la gestión automática de las tareas de bajo nivel. Su sistema de programación es orientado a objetos. La sintaxis es explicita, clara y sencilla, semejante al lenguaje natural, aunado a la disposición de distintos recursos en línea, como herramientas científicas (paquetes) con amplio soporte, como Matplotlib (herramienta para hacer gráficos), numPy y SciPy (aplicaciones numéricas y científicas) y Biopython (herramienta para biología computacional)(Carvajal-Rodríguez, 2015).

### 1.10.2 Software Plink.

Conjunto de herramientas bioinformática de libre acceso con código de abierto de Shaun Purcell y Christopher Chang (Purcell, 2007), cuya función principal es la gestión y el análisis de genotipos con datos tipo SNP, mediante una serie de análisis básicos como la generación de archivos, filtrado de datos, la estadística de la información, y el análisis de asociación del genoma completo. A su vez se encuentra integrado a PLINK, la interfaz gPLINK y Haploview, paquete de software basado en Java, la cual permite la visualización, anotación y almacenamiento de los resultados, de forma sencilla, mediante cuadros de dialogo utilizando comandos de PLINK válidos y comunes.

### 1.10.3 Software R.

Es un programa gratuito integrado por un conjunto de herramientas bioinformáticas de código abierto, desarrollado por Bell Laboratories el cual por medio de una variedad de técnicas estadísticas (modelado lineal y no lineal, pruebas estadísticas clásicas, análisis de series temporales, clasificación, agrupamiento, etc.) y graficas permite el análisis, manipulación, cálculo y visualización de datos (R Development Core Team, 2021).

### 1.10.4 Algoritmo para Detección de Variaciones Estructurales

Las Variaciones estructurales fueron estimadas utilizando la definición de Salomón-Torres et al., (2014), implementada siguiendo el siguiente procedimiento:

1. Obtener (Obtención del) Desequilibrio por Ligamento y el Vector de Medias:

   Paso 1: Ordenar los SNPs por posición, de menor a mayor. Llamar al conjunto: ***Lista de SNPs iniciales.*** 5 kb, (0kb – 5kb; 5kb – 10kb, así sucesivamente hasta el intervalo de 95kb – 100kb) obtener la media

de LD para cada intervalo. Al conjunto de medias resultantes se le denomina: Vector de Medias.

2. Determinar SNPs con LD significante. Seleccionar pares de SNPs con LD significante.

   Paso 4: Considerar la ***Lista de SNPs iniciales*** *y el* ***Vector de Medias****. P*ara cada SNP hacer una *prueba t* de igualdad de medias del LD con cada uno de sus vecinos en 100 en kb, tomando en cuenta la media de su intervalo respectivo. Formar una lista con los ***valores p*** de significancia estadística.

   Paso 5: Aplicar la corrección de Benjamini-Hochberg a la lista de ***valores p***, con lo que se obtiene los valores corregidos (valor q)

   Paso 6: De la ***Lista de SNPs*** *iniciales,* eliminar como su vecino todos los SNPs cuyo *valor q* no sea significante. Llamar al conjunto resultante: ***Lista de SNPs con vecinos significantes*****.**

3. Obtener la lista de SNPs que tengan al menos 15 SNPs vecinos significantes en el rango de 100kb:

   Paso 7: De la ***Lista de SNPs con vecinos significantes*** encontrar todo los SNPs que tengan al menos 15 vecinos significantes en el rango de 100 kb. Llamar a este conjunto: ***Lista de vecinos con 15 significantes***.

4. De la ***Lista de Vecinos con 15 significantes***, obtener el conjunto de SNPs cuyos vecinos en 100kb, todos tiene LD arriba de la media, o, todos tiene LD debajo de la media:

   Paso 8: De la ***Lista de vecinos con 15 significantes***, encontrar los SNPs cuyo LD con todos sus vecinos, son todos mayores a la media de su intervalo respectivo, o todos menores a la media de su intervalo respectivo. Llamar a conjunto de datos resultantes como: *SNPs adecuados*.

5. De la lista de ***SNPs adecuados***, seleccionar grupos al menos tres SNPs adyacentes, cuya longitud del primero al último SNP sea 1 kb o más, y que

el LD de cada SNPs con todos sus vecinos sea consistente; esto es: todos los SNPs tengan solo vecinos arriba de la media, o todos los SNPs tengan solo vecinos por debajo de la media. Llamar a cada región formada por un grupo: ***Variación Estructural de alta densidad basada en LD de rango corto***.

# CAPÍTULO II. ANÁLISIS DE VARIACIONES ESTRUCTURALES BASADOS EN LD DE RANGO CORTO EN 10 POBLACIONES NATIVAS Y MESTIZAS DE MÉXICO.

## 2.1Resumen

La población mexicana es reconocida por la diversidad genética que presenta su población. El objetivo de este análisis es estudiar las variaciones estructurales basadas en LD de rango corto en poblaciones nativas y mestizas de México, y su relación con la región geográfica donde han evolucionado.

La información corresponde a datos genómicos de 10 poblaciones mexicanas, proporcionados por Instituto Nacional de Medicina Genómica (INMEGEN). Esta consta de 785,663 SNPs de 383 individuos de 7 poblaciones mestizas y 3 nativas. Primero se aplicaron filtros de calidad eliminando todos los SNPs que no cumplían con la frecuencia de alelo menor (MAF) <0.05, ni con el principio de Hardy Weinberg. Con los datos filtrados se utilizó la definición propuesta por Salomón-Torres et al., (2014) en donde: para cada raza y cada cromosoma, se calcular el Desequilibrio por Ligamento ($r^2$) de rango corto (≤100 Kb). Se ordenan los pares de SNPs por distancia y se obtiene un conjunto de medias de LD (llamarlo Medias Esperadas) usando intervalos de 5 Kb. Se inspeccionan todos los SNPs, ordenados de menor a mayor, buscando segmentos de al menos 1 Kb, que consistan de al menos tres SNPs adyacentes, de tal forma que: para cada SNP, la $r^2$ con sus vecinos en un rango de 100 Kb, hacia la derecha del SNP, todas sean menor, o todas sean mayor, a su correspondiente media de LD esperada, y su *valor p* sea significante después de una corrección de múltiples pruebas de Benjamini-Hochberg. Por último, se consideran sólo regiones de SNPs distribuidos de forma

homogénea, y, solo SNPs que tengan 15 SNPs vecinos en un rango de 100 Kb. Los segmentos resultantes son llamados Variaciones Estructurales de Alta Densidad basadas en LD de Rango Corto. Se identificando 4,374 VEs en las 10 poblaciones. Se ubicaron a 8,443 genes. El gen FTO resultó consistente en las 10 poblaciones. Mediante un análisis de agrupación se visualizó principalmente la separación de las poblaciones nativas y mestizas, seguido de un agrupamiento de por zona geográfica en la población mestiza. Las VEs permiten identificar el componente genético ancestral principal en las poblaciones mestizas y nativas; y se agrupan por la región geográfica donde han evolucionado las distintas poblaciones.

## 2.2 Introducción

La población mexicana actual está conformada principalmente por dos grupos: mestizos y nativos. La población mestiza se originó por varios acontecimientos históricos, como la conquista de América por los españoles, y la llegada al territorio mexicano de grupos pobladores de África y Asia, entre otros. Estas influencias extranjeras originaron mestizaje con los diferentes grupos étnicos en el México precolombino, produciendo una amplia diversidad poblacional.

Por lo contrario, la población nativa corresponde a pueblos originarios de México, que, no se han mezclado durante siglos debido a sus propias ideologías culturales, y a su distancia geográfica con los asentamientos urbanos de las poblaciones mestizas modernas. Estas poblaciones nativas han reduciendo su tamaño por distintos motivos, entre ellos los eventos epidemiológicos de viruela, sarampión y tifoidea sucedidos entre 1521 y 1580 que provocaron una disminución considerable de las poblaciones nativas. A su vez, las distintas poblaciones nativas

se establecieron en ciertas regiones específicas del territorio (mayormente centro y el sureste del país), aprovechando la fertilidad de la tierra para la producción de sus cultivos de maíz, frijol y cacao; adaptándose y sobreviviendo a las distintas condiciones presentes en su entorno (Martínez-Cortés et al., 2012; Rubi-Castellanos et al., 2009).

El censo de población 2020 en México, realizado por el Instituto Nacional de Estadística y Geografía (INEGI), mostró que la población total mexicana, en 2020, fue de 126,014,026 personas. Más del 78% eran mestizas; el 19.4% pertenecía a grupos poblacionales indígenas (nativos), y el 2% estaba representado por personas afrodescendientes(Instituto Nacional de Estadística y Geografía, 2022).

En la última década, en México, estudios demográficos, complementados con estudios de genética de poblaciones, han revelado un alto grado de diversidad genética, con una distribución desigual de alelos, que varía de acuerdo a la región geográfica analizada(Aguilar-Ordoñez et al., 2021; Ávila-Arcos et al., 2020; Moreno-Estrada et al., 2014; Romero-Hidalgo et al., 2017). Estos estudios demuestran que la evolución de la población mexicana ha sido modelada por una serie de eventos históricos, epidemiológicos, sociales, culturales, demográficos y económicos.

Uno de los primeros estudios sobre diversidad genética en poblaciones latinoamericanas fue realizado por Silva-Zolezzi et al., (2009). El estudio incluyó 300 mestizos no relacionados, de siete poblaciones geográficamente separadas de México y 30 nativos de la etnia Zapoteca. Un conjunto de 100k genotipos fue muestreado en cada individuo, analizando la diversidad genética, patrones de desequilibrio por ligamento y la cantidad de haplotipos compartidos. Los resultados

arrojaron las primeras mediciones de la diversidad genética, que pueden ser consideradas para estudios de asociación de enfermedades complejas en la población mexicana. Por su parte, Martínez-Cortés et al., (2012) en su estudio genético de diversidad poblacional, encontró que la ascendencia paterna de la población mestiza mexicana está conformada por 64.4% de ancestría Europea, el 30.8% de ancestría nativa americana y en un mínimo porcentaje de ancestría africana y asiática.

En el estudio reportado por Moreno-Estrada et al., (2014), se analizaron cerca de 1 millón de polimorfismos de nucleótido simple (SNP), en el genoma completo de 1000 individuos pertenecientes a 20 poblaciones indígenas, y 11 poblaciones mestizas de México, identificando variaciones genómicas que aportan información sobre la estratificación geográfica de las variaciones genéticas presentes en la población mexicana, así como su futura aplicación para habilitar la medicina de precisión. En tanto, en el estudio realizado por Romero-Hidalgo et al., (2017), analizaron secuencias de genoma completo para reconstruir la historia demográfica de 12 genomas de Nativos americanos, pertenecientes a seis distintos grupos étnicos, que representan los tres clústeres principales de México (del Norte, del Sur y Mayas). El estudio mostró que la diversidad de poblaciones de Nativos Americanos, particularmente las poblaciones meridionales, es mayor de lo que se había descrito previamente por Moreno-Estrada et al., (2014). Aunado a que los procesos demográficos y adaptativos que sufrieron estos grupos moldeó su arquitectura genética, teniendo implicaciones en procesos biológicos relacionados con la salud de la población actual de Nativos y mestizos. Mientras en el estudio de

variaciones utilizando el genoma completo en 27 poblaciones indígenas de México, realizado por Aguilar-Ordoñez et al., (2021), reportó un total de 497 variaciones comunes de nucleótido simple (SNV, por sus siglas en ingles), localizadas en genes relacionados con respuesta a fármacos, más 316,577 dentro de promotores y potenciadores de la trascripción de genes. Algunas de estas variaciones están en el gen PPARG, que es un receptor nuclear asociado con problemas de salud altamente prevalentes en la población mexicana, tal como la obesidad, diabetes y la resistencia a la insulina.

La población mexicana comparte diversas características, como el territorio, costumbres y lengua. No obstante, existen diferencias genéticas dentro del mismo territorio siendo contrastantes las características físicas y genéticas. Por ello, recientemente se han estudiado la estructura genética de las poblaciones mestizas y nativas de México; para poder establecer patrones de variaciones genéticas de ancestría nativa o derivadas del mestizaje, que pueden estar asociadas a enfermedades complejas; y características físicas y fisiológicas. El análisis de la variabilidad genética permite definir la diversidad de la población mexicana y su impacto en la salud.

Se reconocen como variaciones estructurales (VEs) el reordenamiento de ADN que influye en la configuración de la estructura y función del genoma. Las VEs por si solas o aunadas a factores ambientales pueden ser causa de expresiones fenotípicas o de la presencia de enfermedades genéticas, impactando en la salud de los individuos. Por ejemplo, existen VEs que están asociadas con ciertas

condiciones o patologías, como el cáncer, discapacidad cognitiva y la predisposición a la obesidad (Sudmant et al., 2015).

En esta investigación se presenta la identificación de VEs bajo la definición establecida por Salomón, 2014. Dichas VEs están basadas en el desequilibrio por ligamiento (LD) de rango corto. Para ello es necesario tomar en cuenta regiones de SNPs, que contengan al menos 15 SNPs adyacentes; calcular el desequilibrio ligado ($r^2$) de corto alcance (≤100 Kb), para cada una de las poblaciones y sus cromosomas; obteniendo las medias de las LD y buscar en segmentos de al menos 1Kb, grupos de al menos 3 SNPs adyacentes que sus $r^2$ , estén todas por arriba o por debajo de la media esperada y sus valores P de la prueba-t, para igualdad de medias sean significativas después de la prueba de corrección de Benjamini-Hochberg (Salomón-Torres et al., 2014).

Utilizando los datos genómicos, proporcionados por INMEGEN, de 10 poblaciones mexicanas, que corresponden a 7 poblaciones mestizas geográficamente distantes distribuidas en las regiones del norte, Sonora y Zacatecas; Noreste, Tamaulipas; Centro-Sur, Guanajuato y Guerrero; sureste, Veracruz y Yucatán. Además de poblaciones 3 nativas que abarcan las siguientes regiones, entre los estados del norte de Durango, Zacatecas, Nayarit y sur de Chihuahua, la población de los Tepehuanos; en el estado de Oaxaca, región de la Sierra, Valles Centrales y el Istmo de Tehuantepec, la población Zapoteca, por último, la población Maya ubicada en los estados de Yucatán, Campeche y Quintana Roo (Instituto Nacional de Pueblos Indígenas, n.d.; Moreno-Estrada et al., 2014; Sistema de informacion Cultural, 2023).

En esta investigación se utiliza un nuevo enfoque para la identificación de VEs basadas en LD, con ella se logró identificar un total de 4374 VEs en las 10 poblaciones, describiendo una distribución claramente diferenciada entre las poblaciones nativas y mestizas y entre cada una de ellas, considerando la estructura genética y la demografía de las poblaciones.

## 2.3 Materiales y Métodos.

### 2.3.1 Declaración de Ética para la obtención de datos.

Este estudio se realizó en conformidad con la Declaración de Helsinki y la aprobación del Comité de Investigación, Ética y Bioseguridad del INMEGEN. A su vez con el apoyo de la Comisión Nacional para el Desarrollo de Pueblos Indígenas. Con la aceptación de cada uno de los participantes, firmando de manera autógrafa o por medio de huella digital el consentimiento informado.

### 2.3.2 Descripción de los datos.

El conjunto de datos fue proporcionado por el Instituto Nacional de Medicina Genómica del gobierno de México (http://inmegen.gob.mx). La muestra consistió en haplotipos de 785,663 polimorfismos de nucleótido simple (SNP), de 383 individuos mexicanos de 10 poblaciones. De las poblaciones, tres fueron de nativos y siete de mestizos. La distribución de poblaciones e individuos es la siguiente: Las poblaciones nativas fueron Maya, con 30 individuos, Tepehuano, con 20 individuos y Zapoteca, con 21 individuos. Las poblaciones mestizas fueron: Guanajuato, con 48 individuos, Guerrero, con 50 individuos, Sonora, con 48 individuos, Tamaulipas, con 17 individuos, Veracruz, con 50 individuos, Yucatán, con 49 individuos y Zacatecas, con 50 individuos.

### 2.3.3 Control de calidad de los datos.

Para tener datos libres de errores se aplicaron filtros de control de calidad. Se eliminaron todos los SNPs cuya Frecuencia de Alelo Menor (MAF) fue < 0.05. También se eliminaron todos los SNPs que no cumplieron con el principio de Hardy-Weinberg (< 0.0001). La cantidad inicial de SNPs fue de 7, 856,630, y después de aplicar el control de calidad resultaron 6,197,471 SNPs. El promedio de SNPs por población, después del control de calidad fue de 619,535 SNPs. Con este último conjunto se realizaron los análisis de este trabajo.

### 2.3.4 Desequilibrio por Ligamento.

El desequilibrio por ligamento de todos los pares de SNPs de cada cromosoma, dentro de cada población, fue estimado mediante la fórmula de correlación de Pearson ($r^2$):

$$r^2 = \frac{(p_{11} - p_1 q_1)^2}{(p_1 q_1 p_2 q_2)}$$

Donde $p_1$ y $p_2$ son las frecuencias de alelo menor y alelo mayor, respectivamente, del SNP1, y, $q_1$ y $q_2$ son las frecuencias de alelo menor y alelo mayor, respectivamente, del SNP2. $p_{11}$ corresponde a la frecuencia de observar ambos alelos menores en el mismo individuo en toda la población.

Para evitar errores por el tamaño de la muestra de cada población, se aplicó la siguiente corrección al desequilibrio por ligamento:

$$r^2 corrected = \frac{r^2 computed - \frac{1}{n}}{1 - \frac{1}{n}}$$

Donde *n* corresponde al número de haplotipos en la muestra (Villa-Angulo et al., 2009). Los valores de $r^2$ fueron estimados con la herramienta PLINK(Purcell, 2007).

### 2.3.5 Variaciones Estructurales de Alta Densidad basadas en LD de Rango Corto.

Para inspeccionar patrones de LD a lo largo de todo el genoma, implementamos la definición de Variación Estructural basada en LD de Rango Corto hecha por Salomón-Torres et al., (2014). Esto se describe de la siguiente manera:

Para cada raza y cada cromosoma, calcular el Desequilibrio por Ligamento ($r^2$) de rango corto (≤100 Kb). Ordenar los pares de SNPs por distancia y obtener un conjunto de medias de LD (llamarlo Medias Esperadas) usando intervalos de 5Kb. Inspeccionar todos los SNPs, ordenados de menor a mayor, buscando segmentos de al menos 1 Kb, que consistan de al menos tres SNPs adyacentes, y que, para cada SNP, la $r^2$ con sus vecinos en un rango de 100 Kb, hacia la derecha del SNP, todas sean menor, o todas sean mayor, a su correspondiente media de LD esperada, y su *valor p* sea significante después de una corrección de múltiples pruebas de Benjamini-Hochberg. Por último, para considerar solo regiones de SNPs distribuidos de forma homogénea, considerar solo SNPs que tengan 15 SNPs vecinos en un rango de 100 Kb. Llamar a estos segmentos: Variaciones Estructurales de Alta Densidad basadas en LD de Rango Corto.

### 2.3.6 Corrección de Múltiples Pruebas

Al conjunto de valores *p* resultante de las pruebas de igualdad de medias del LD de cada SNP con sus vecinos en un rango de 100 kb, se le aplicó la corrección de múltiples pruebas de Benjamni-Hochberg, con el fin de controlar la tasa de falsos positivos. El procedimiento fue el siguiente: se ordenó el conjunto de *valores p* de menor a mayor; definiendo desde el primero hasta el último *valor p*, como el i-esimo valor p, *p(i)*. En seguida, se inspeccionó la lista de valores *p*, de mayor a menor, haciendo el siguiente calculo: $\frac{i}{m}$, donde *i* es el índice del *valor p,* y *m* es el total de valores *p*; y se comparó con el *i*-esimo *valor p. Se hizo este* cálculo hasta que el *i*-esimo *valor p* fue menor o igual a $0.05 * \frac{i}{m}$. Se declararon todos los valores *p* restantes, hasta el más pequeño, como significantes (Benjamini et al., 2001).

### 2.3.7 Análisis de Componentes Principales.

Con el total de Variaciones Estructurales encontradas por cada cromosoma, en cada población, se generaron vectores de 22 dimensiones y se aplicó un Análisis de Componente Principales (PCA) buscando cúmulos que permitieran clasificar los resultados. PCA es una técnica para reducir la dimensionalidad de un conjunto de datos, que posee un gran número de variables interrelacionadas, al tiempo que retiene lo más posible de la variación presente en el conjunto de datos. PCA genera un nuevo conjunto de variables, los componentes principales (PCs), los cuales son ortogonales y no están correlacionados, ordenándolos de tal manera que los primeros retienen la mayor variación presente en todas las variables originales (Jolliffe, 2002).PCA es en teoría, la transformación óptima para un conjunto de datos, en términos de mínimos cuadrados. El procedimiento para obtener los

componentes principales puede resumirse de la siguiente manera: dado un vector $X^T$ de $n$ dimensiones, $X = [x_1, x_2 \ldots, x_n]^T$, del cual sus vectores de medias, $M$, y covarianzas, $C$, son descritos por: $M=E(X)=[m_1, m_2, \ldots, m_n]^T$ y $C=E[(X-M)(X-M)^T]$; calcular los eigenvalores $\lambda_1, \lambda_2, \ldots, \lambda_n$, y los eigenvectores $P_1, P_2, \ldots, P_n$; y ordenarlos de acuerdo a su magnitud $\lambda_1 \geq \lambda_2 \geq \ldots \geq \lambda_n$. Seleccionar $d$ eigenvectores para representar las $n$ variables, $d<n$. Entonces $P_1, P_2, \ldots, P_d$ son llamados componentes principales. Se realizó el análisis de componentes principales (PCA) utilizando el software R (R Development Core Team, 2021).

## 2.4 Resultados.

### 2.4.1 Distribución de frecuencia de alelo menor.

Para investigar que tan informativos son los SNPs muestreados en las diez poblaciones, estimamos la distribución de las frecuencias alélicas (la Figura 5 muestra las proporciones de los grupos: Nativos, Mestizo área costera y Mestizo área no costera, y el promedio de la frecuencia de alelo menor (MAF) del conjunto completo de SNPs de cada población. Las tres poblaciones nativas, Maya, Tepehuano y Zapoteca presentaron las proporciones polimórficas más bajas, entre 71% y 83% comparado con las poblaciones restantes (Nativas), las cuales presentaron entre 88% y 96%. Por lo que, una fracción sustancial de SNPs son informativos en todas las poblaciones del estudio. La Figura 5 presenta todos los SNPs (incluyendo SNPs monomórficos y polimórficos), sin embargo, los SNPs monomórficos fueron excluidos para todos los análisis subsecuentes.

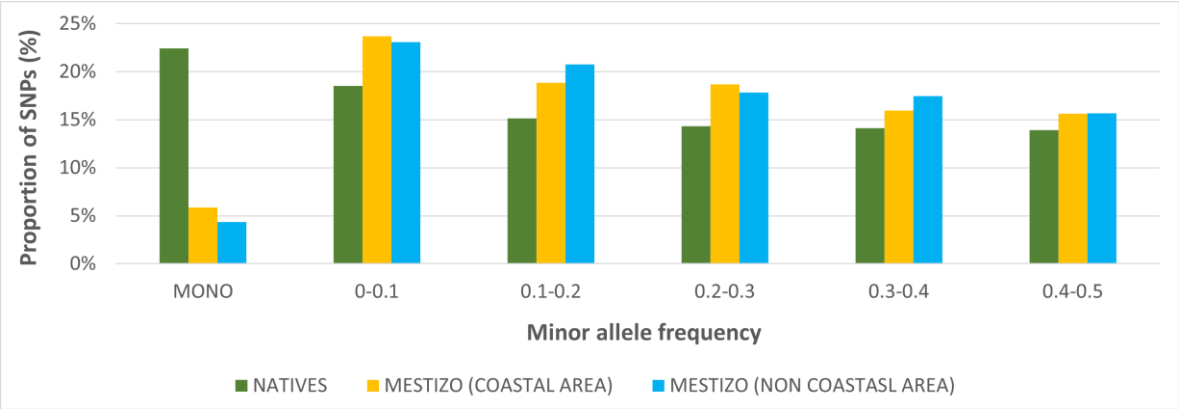


**Figura 5**. Distribución de MAF. Promedio de las proporciones de las frecuencias de SNPs por grupo poblacional.

En general, las tres poblaciones nativas exhiben los valores MAF más bajos. Esto puede deberse a que los SNPs muestreados en el ensayo de genotipificación fueron descubiertos inicialmente en poblaciones mestizas. Los valores MAF varían desde un máximo de 0.227 (Sonora) hasta 0.177 (Zapotecas), lo cual es una diferencia del 10% en la escala completa de 0.0 a 0.5. La disminución promedio del MAF entre las poblaciones es de 1.1%.

### 3.4.2 Análisis de LD.

El conjunto de 619,535 SNPs promedio por población, obtenido después del control de calidad, fue utilizado para evaluar la medida de LD entre pares como función de la distancia física, y para estimar la población efectiva en 10,000 generaciones previas, para cada población.

El coeficiente de correlación de LD, $r^2$, fue calculado de los haplotipos de todos los pares de SNPs de cada cromosoma, para toda la población. La figura 6 muestra el promedio de los valores de $r^2$ usando rangos de 5 kb. La caída de LD en función de la distancia es rápida, de tal manera que, $r^2$ promedia ~0.10 en 100 kb.

Las poblaciones nativas presentan uniformemente valores más altos de LD, en relación a las otras poblaciones. En distancias más cortas (5 kb), los zapotecas, tepehuanos y mayas presentan los valores más altos de $r^2$ (~0.49 en promedio), mientras que el resto de las poblaciones presenta un valor promedio de ~0.37. En distancias más largas (100 kb), los zapotecas y tepehuanos presentan consistentemente los valores más altos, con un promedio de ~0.156. Las poblaciones de Guanajuato, Guerrero, Sonora, Veracruz, Yucatán y Zacatecas presentan los valores más bajos, con un valor promedio de 0.075. Mientras que las poblaciones Maya y Tamaulipas presentan valores intermedios, con un valor promedio de 0.115.

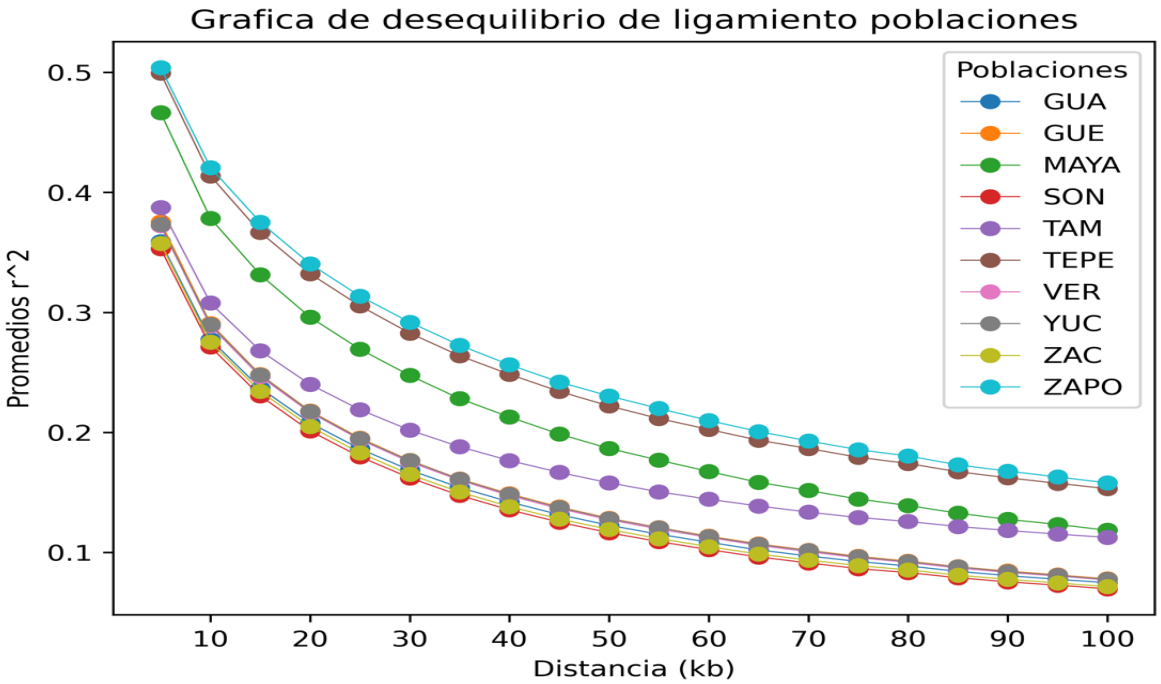


**Figura 6.** Desequilibrio de Ligamiento. El LD muestra un rápido descenso en relación con el promedio de las $r^2$ ~0.10 a 100kb. Se utilizan rangos de 5kb para el promedio de las $r^2$.

### 2.4.3 Variaciones Estructurales Basadas en LD de Rango Corto.

Variaciones Estructurales basadas en LD de rango corto (≤100 kb) fueron estimadas usando la definición de Salomón-Torres et al.,( 2014), presentada en la sección de Métodos y Métodos. En la tabla 1, se detalla las características de las Variaciones Estructurales en todas las poblaciones.

**Tabla 1**. Estadística general de Variaciones Estructurales.

| Población | SVs | No. SNPs | SNPs max | Promedio SNP | Promedio por cromosoma SVs | Tamaño total (Kb) | Tamaño max. (Kb) | Tamaño Min.(Kb) | Promedio del tamaño (Kb) | Promedio del tamaño del SVs 95% Intervalo de Confidencia (min, max Kb) | |
|---|---|---|---|---|---|---|---|---|---|---|---|
| **GUA** | 325 | 21805 | 694 | 67.09 | 14.77 | 2486957.14 | 96202.84 | 3.37 | 7652.18 | 6502.98 | 8801.37 |
| **GUE** | 351 | 22955 | 475 | 65.40 | 15.95 | 2468700.37 | 57122.29 | 1.48 | 7033.33 | 6086.68 | 7979.99 |
| **MAYA** | 582 | 18602 | 424 | 31.96 | 26.45 | 2275563.84 | 50670.71 | 1.18 | 3909.90 | 3436.35 | 4383.46 |
| **SON** | 362 | 23086 | 654 | 63.77 | 16.45 | 2509560.57 | 65696.73 | 1.73 | 6932.49 | 5981.81 | 7883.17 |
| **TAM** | 288 | 10135 | 251 | 35.19 | 13.09 | 2490069.32 | 82406.69 | 5.38 | 8646.07 | 7417.59 | 9874.56 |
| **TEPE** | 678 | 14639 | 200 | 21.59 | 30.82 | 2217850.07 | 50505.33 | 1.34 | 3271.17 | 2875.86 | 3666.47 |
| **VER** | 354 | 23359 | 634 | 65.99 | 16.09 | 2487290.10 | 80124.25 | 1.56 | 7026.24 | 6047.85 | 8004.64 |
| **YUC** | 363 | 23082 | 470 | 63.59 | 16.50 | 2519096.64 | 49639.78 | 1.24 | 6939.66 | 6058.78 | 7820.54 |
| **ZAC** | 342 | 23503 | 507 | 68.72 | 15.55 | 2515444.93 | 58965.44 | 1.41 | 7355.10 | 6386.89 | 8323.31 |
| **ZAPO** | 730 | 13210 | 154 | 18.10 | 33.18 | 2076801.41 | 36624.85 | 1.12 | 2844.93 | 2533.07 | 3156.79 |

En la tabla 1, se muestran las estadísticas de las variaciones estructurales totales de las poblaciones analizadas. Se puede observar que las poblaciones con mayor cantidad de Ves, identificados son las poblaciones nativas, siendo la población Zapoteca la que presenta un mayor valor, con 733 VEs. Por su parte, de las poblaciones mestizas la que obtuvo mayor cantidad de VEs es la población de Yucatán con 364. En contraste la población con menor número de VEs identificados fue la población de Tamaulipas con 288 VEs.

En relación al tamaño de las VEs identificadas, el promedio total es de 6,161 kb, siendo la población de Guanajuato la que presenta un mayor valor de distancia de

VEs con 96,202.8 Kb, mientras la población de Zapotecas presenta el menor valor de la distancia con menor valor con 1.12 Kb.

Dentro de las VEs se identificaron 8443 genes. Del total de genes identificados, 6293 se encuentran en más de una población. De esta cantidad, 790 los comparten en las 10 poblaciones, mientras 2190 son genes que solamente se localizan en una de las 10 poblaciones.

Considerando las posiciones de nuestras VEs identificadas, se detectaron 506 regiones. Ninguna de las regiones es compartida entre el número total de poblaciones. 1353 regiones son compartidas entre 2 o más poblaciones, mientras solamente una región es compartida por 6 de las 10 poblaciones (Guerrero, Sonora, Tamaulipas, Tepehuanas, Veracruz y Zacatecas).

Con la información del VEs obtenidas por cromosoma y por población se realizó un análisis de PCA. Primero, se utilizaron solamente VEs identificadas debajo de la media. Segundo, se utilizaron solo con VEs identificados por encima de la media. Y, el último con el total de VEs encontradas.

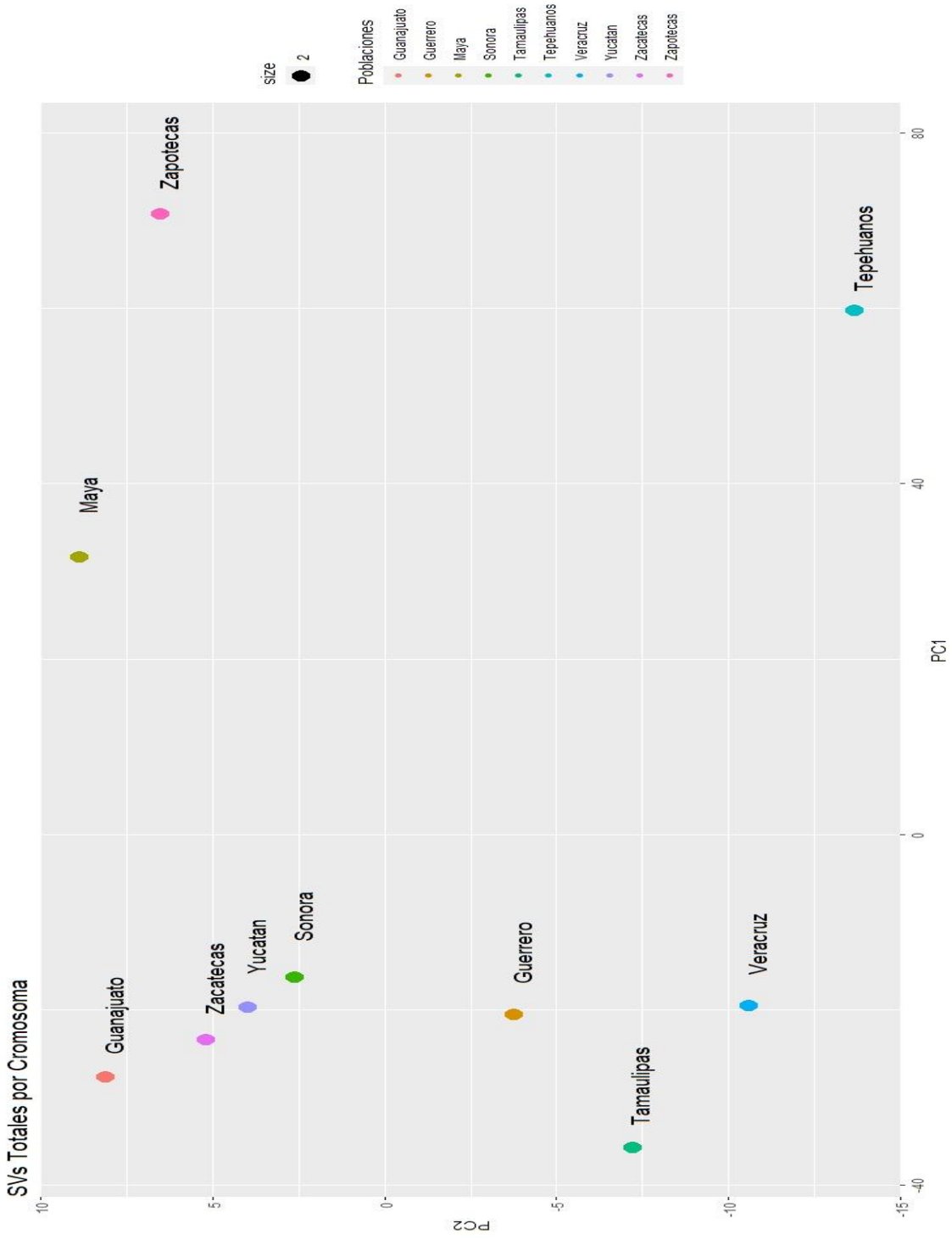


**Figura 7**. PCA de VEs totales por cromosoma identificadas en cada población.

En la figura 7 se observa la gráfica de PC1 vs PC2 utilizando el total de VEs. Las poblaciones nativas, zapotecos, mayas y tepehuanos aparecen en la región de

valores positivos de PC1, el resto de poblaciones, que son mestizas se presentan en los valores negativos de PC1.

Las poblaciones mestizas, en el PCA de VEs totales forman dos subgrupos, el subgrupo A ubicado en los valores negativos de PC2, formado por las poblaciones de Guerrero, Tamaulipas y Veracruz; y el grupo B, ubicado en los valores positivos de PC2, formado por Guanajuato, Zacatecas, Yucatán y Sonora. Considerando la distribución geográfica de nuestro país, en zona norte, centro, sur y sureste, las poblaciones del subgrupo A, lo conforman poblaciones geográficamente localizadas en la región del sur, centro y norte del país. Mientras el subgrupo B, lo conforman principalmente poblaciones de la región del centro, norte, y sureste.

## 2.5 Discusión

En este estudio se realizó la búsqueda e identificación de variaciones estructurales (VEs) basadas en LD de rango corto en 10 poblaciones mexicanas. El LD puede ser influenciado por la estructura de la población y la historia demográfica, como lo menciona, Ávila-Arcos et al., (2020). El método de identificación de VEs basadas en LD de rango corto, se introdujo anteriormente para detectar variaciones estructurales en ganado bovino, mostrando la relación de las razas por su región geográfica y su evolución, resultado del número de variaciones estructurales obtenidas por cada raza (Salomón-Torres et al., 2014).

En nuestro análisis identificamos 4,374 VEs basadas en LD de rango corto, en poblaciones humanas mestizas y nativas de México. En estudios previos sobre la identificación de patrones y de diversidad de estructuras genéticas en la población mestiza y nativa, se han empleado diferentes métodos y tecnologías. Ávila-Arcos et

al., (2020), Aguilar-Ordoñez et al., (2021), reportaron 120,735 y 8.68 millones de SNVs respectivamente, mientras Romero-Hidalgo et al., (2017) 332,272 SNVs, en poblaciones nativas. Por su parte Moreno-Estrada et al., (2014) y Silva-Zolezzi et al., (2009), mediante análisis de heterocigocidad (HET) y $F_{ST}$ pareado, de la población mestiza y nativa reportaron la diversidad en la estructura genética de la población mexicana.

Con las VEs resultantes de nuestro análisis de la población mexicana se efectuó un análisis de PCA, en donde la distribución de la información sugiere, la ubicación geográfica de las poblaciones, el aislamiento de los grupos nativos así como la migración de estos grupos a otras regiones, y las proporciones de mezcla entre las mismas poblaciones o con otras poblaciones foráneas.

Los resultados obtenidos en los análisis de proporciones polimórficas, de desequilibrio ligado, así como en el PCA producido con el número total de VEs identificadas en cada cromosoma de cada población, son congruentes, reflejando una clara diferencia entre las poblaciones mestizas y nativas, entre ellas como dentro de cada una de las poblaciones nativas.

Las poblaciones nativas mostraron su particularidad separándose cada una de ellas, como lo mencionan Moreno-Estrada et al., (2014), en su análisis con 20 grupos indígenas en donde los grupos nativos mexicanos se diferencian entre ellos, regionalizándose según su ubicación geográfica en tres zonas principalmente, los del norte considerando a los tepehuanos, los del sur a los zapotecos y los mayas. Aunado a esto, Romero-Hidalgo et al., (2017), menciona que los grupos nativos están conformados por distinto componente ancestral, el componente del norte, el

componente del sur y el componente maya; en donde la población tepehuana muestra un valor mayor de proporción del componente del norte, mientras la población zapoteca a la del sur y los mayas a la propia.

De las tres poblaciones nativas la población zapoteca presentó una mayor proporción de SNPs monomórficos, además de una caída de LD más lenta y el mayor número de variaciones estructurales. Estos resultados sugieren que el grupo nativo zapoteca corresponde a una población con menor intercambio genético con otras poblaciones, producto de un mayor aislamiento propiciado principalmente de las barreras geográficas y culturales. Sin embargo, debemos de considerar la posibilidad que cierta cantidad SNPs muestreados no son polimórficos para esta población, mientras si lo son para las otras.

En relación a las poblaciones mestizas nuestro estudio muestra en relación a los resultados del PCA derivado de las variaciones estructurales totales por cromosoma identificadas en cada población, se conforman 2 subgrupos, ubicando a las poblaciones de Guerrero, Tamaulipas y Veracruz en un grupo, mientras Guanajuato, Zacatecas, Yucatán y Sonora en otro.

En el grupo A, la población de Tamaulipas, Veracruz y Guerrero se encuentran en la zona centro-costera de México. El grupo B, los estados de Guanajuato y Zacatecas se encuentran ubicados en la zona centro del país. Seguidos de Yucatán y Sonora, los cuales no se encuentran geográficamente cerca. La distribución que se presenta puede ser resultado de varios factores como, la ubicación geográfica, el porcentaje de contribución genética derivado de las mezclas de poblaciones extranjeras, la convergencia genética o a la proporción de

componente ancestral específico para cada región, el componente del norte, el componente del sur y el componente maya como lo mencionan Moreno-Estrada et al., (2014) y Romero-Hidalgo et al., (2017). Sumando a lo anterior Raghavan et al., (2015), establecen la existencia de la continua genética de ancestría presente en las poblaciones actuales, los cuales anteceden por lo menos de 8,500 años. Por otra parte, estudios antropológicos realizados por González et al., (2003), indican que los primeros vestigios humanos en México se presentaron aproximadamente hace 12,000 años, en la zona centro de México y en la península de Yucatán. Mostrando con ello una relación entre las poblaciones mestizas presentes en este estudio.

Con la información obtenida de los VEs en las 10 poblaciones, se examinaron bases de datos (NCBI) en la búsqueda de los genes involucrados en nuestras variaciones estructurales reconocidas, identificando genes como el FTO y el ABCA1 en las 10 poblaciones. Estos genes han sido relacionados con la obesidad, en la función del tejido adiposo, así como en la distribución de la grasa en la población mexicana, como lo mencionan distintos estudios (Chama-Avilés et al., 2023; Villalobos-Comparán et al., 2017). A su vez el gen ELMO1, el cual esta sugiere la susceptibilidad a la nefropatía diabética y la diabetes tipo II, como lo menciona Aguilar-Ordoñez et al., (2021).

Sumado a esto, se identificaron los genes BRCA2, en las poblaciones de Guerrero, Maya y Zapoteca, así como el gen e IKBKB en Guanajuato, Sonora, Tamaulipas, Tepehuanos y Yucatán. Estos genes fueron mencionados anteriormente por Moreno-Estrada et al., (2014) y Aguilar-Ordoñez et al., (2021)

asociándolos con el desarrollo de cáncer de mama y de ovario en la población mexicana.

Es importante mencionar la detección de genes reportados en la población mexicana por distintos autores(Aguilar-Ordoñez et al., 2021; Ramírez-García et al., 2013), aunado a los genes anteriormente mencionados en los que son compartidos por las 10 poblaciones, también se encontraron genes que solamente se localizaron en una de las poblaciones, como es el gen MCHR1, implicado con la regulación neuronal del consumo de alimentos, el cual se ubicó solamente en la población de Zacatecas. Mientras que el gen SLC30A8, se identificó solamente en la población de Guanajuato, este gen confiere cierta disposición a la diabetes mellitus no insulino-dependiente. Así como el gen IGF2BP2, localizado solo en la población de Tamaulipas, el cual juega un rol importante con el metabolismo, asociado a la susceptibilidad de diabetes.

Por otra parte, también se encontraron genes relacionados con enfermedades degenerativas de la columna y el incremento de riesgo del desarrollo de hernia de disco lumbar, como el gen IL-6 en la población de Veracruz y el IL-10 y el CILP presentes en la población Zapoteca, estos genes intervienen directamente en la citoplasma del núcleo pulposo del disco intervertebral, lo que pudiera alterar la composición del mismo (Liu et al., 2022; Yewen Guan, Siting Wang, Jiaqi Wang, Dihua Meng, Huihong Wu, 2022) provocando la aparición de la hernia de disco lumbar. Mientras los genes BDNFs, relacionados con la regulación de la respuesta ante el estrés y con un aumento de riesgo de hernia de disco lumbar en la población China mayor de 50 años se encontraron en la población Maya (Zhu et al., 2018).

La identificación de las SVs basadas en LD de rango corto, que se utilizó en este análisis, muestran patrones de la estructura genética basadas en las condiciones de las poblaciones mestizas y nativas, su desarrollo a través de los años y los distintos factores que afectan o benefician su estructura genética de la población mexicana, reflejándose en características relacionadas con la salud, propias de cada una de los grupos analizados.

# CAPÍTULO III. FACTORES DE RIESGO Y ASOCIACIÓN DE COMPONENTES DE LA COMPOSICIÓN CORPORAL PARA LA HERNIA DE DISCO LUMBAR EN EL NOROESTE DE MÉXICO.

## 3.1Resumen

El objetivo de este análisis fue investigar la asociación de los componentes de la composición corporal con la presencia de hernia de disco lumbar (HDL) y elucidar si alguno de los componentes es un factor de riesgo para el desarrollo del padecimiento, en la población del noroeste de México.

La población estaba compuesta por 90 adultos participantes de un programa de actividad física para personas con sobrepeso y obesidad. De estos, 19 personas estaban diagnosticadas por medio de resonancia magnética con hernia de disco lumbar. La composición corporal se determinó utilizando un analizador de Bioimpedancia eléctrica. Además se aplicaron cuestionarios de Oswetry e IPAQ para obtener mayor información. También se incluyeron mediciones de glucosa, la talla y el historial médico. Por medio de la estadística descriptiva y el análisis de los componentes principales se estableció la estructura de la información permitiendo visualizar los cúmulos generados. Seguido de ello se aplicó el análisis de regresión lineal, encontrando la asociación entre algunas de las variables de la composición corporal con la HDL. Finalmente, se aplicó un análisis de detección de factores de riesgo, Odds ratio, para declarar las variables como factores de riesgo en el desarrollo de hernia de disco lumbar.

Resultando como variables asociadas, el Grado de Obesidad, el Índice de Masa Corporal, Área de Grasa Visceral y la Circunferencia Abdominal (Valores *P* de 0.0388,0.0171, 0.0055 y 0.0032 respectivamente). Después de aplicar el criterio

de Odd Ratio, el Área de Grasa Visceral y la Circunferencia Abdominal fueron declaradas como factores de riesgo para el desarrollo de Hernia de Disco Lumbar en la población de Baja California. Nuestros resultados proveen nueva evidencia para futuros estudios, así como para la prescripción de la actividad física y programas nutricionales que permitan incrementar la musculatura de la zona del core y disminuir el tejido adiposo, para prevenir o corregir el desarrollo de la Hernia de Disco Lumbar en población de Baja California.

### 3.2 Introducción

El sobrepeso y obesidad son un tema de salud pública mundial. Estos son causantes de muchos padecimientos actuales y han sido asociados como factor de riesgo de muchos otros (Must et al., 1999). La hernia de disco lumbar (HDL) es un padecimiento al que se ha asociado como factor de riesgo el sobrepeso y la obesidad, junto con factores genéticos, ambientales y los relacionados con el estilo de vida de las personas. La HDL es una afección musculo esquelética que se produce cuando parte del disco intervertebral (núcleo pulposo) se desplaza hacia la raíz nerviosa, presionándola y produciéndole lesiones neurológicas (Manelfe, 1992). Su relevancia se debe a que es la segunda causa de incapacidad laboral en los Estados Unidos de América, y durante su evolución puede incidir de manera negativa en la calidad de vida del paciente(Ricci et al., 2006).

Estudios como el de TS Han et al.,(1997) han demostrado que la prevalencia del dolor de espalda baja y la hernia de disco lumbar aumentan cuando la población general presenta incrementos de la circunferencia abdominal (CA), la relación

cintura-cadera y el Índice de Masa Corporal (IMC), tanto en hombres como mujeres. El IMC y CA han sido utilizados como métodos para evaluar el sobrepeso y obesidad. Los valores elevados de estas mediciones están referidos como factores predisponentes a la hernia de disco en la región lumbar. También se ha sugerido que factores metabólicos asociados con la obesidad pueden determinar una HDL (Aro & Leino, 1985). En un estudio realizado por Takatalo et al.,( 2013) en Finlandia, sobre la asociación de la obesidad abdominal y la degeneración discal lumbar, en el cual utilizó resonancia magnética y bioimpedancia eléctrica como métodos de medición de la composición corporal, concluyeron que las mediciones de obesidad abdominal y la circunferencia de cintura están asociadas con la degeneración discal en personas adultas.

México es un país con índices elevados de sobrepeso y obesidad (Salud Pública, 2013). En especial en el estado de Baja California el 31.9% de la población adulta (mayor a 20 años) presenta sobrepeso, mientras que el 39.5% es obesa (datos de la Encuesta Nacional de Salud y Nutrición ENSANUT en 2012). Un estudio realizado por Zonana-Nacach et al., (2013), en el sistema de Seguridad Social de municipios de Baja California, se reportó una incidencia de HDL de 46% en personas con edad promedio de 41.8 años, siendo las mujeres las más afectadas, principalmente con hernia de disco en la lumbar 4 y 5.

En este estudio se presenta el procedimiento y los resultados obtenidos del muestreo realizado en Mexicali, Baja California, México, con el fin de determinar si en pacientes con hernia de disco en la región lumbar existe una asociación entre el padecimiento y componentes de la composición corporal, y si alguno de ellos es factor de riesgo para el desarrollo de HDL. Se tomaron mediciones con el método

de bioimpedancia eléctrica, el historial de salud, y se aplicaron cuestionarios a personas diagnosticadas con HDL. Análisis estadístico descriptivo y de Componentes Principales permitieron analizar la estructura de la información y visualizar cúmulos. Un análisis de Regresión Logística permitió encontrar la asociación entre algunas de las variables de composición corporal y la HDL. Finalmente, la aplicación de la Razón Odd permitió declarar algunas de las variables como factores de riesgo para el desarrollo de HDL en la población de Mexicali, Baja California.

## 3.3 Materiales y Métodos

### 3.3.1 Población de estudios.

Se muestrearon 90 personas adultas, de las cuales 78 eran mujeres y 12 hombres, con una edad promedio de 57.4 años y una desviación estándar de 11.51. Todos los participantes acudían a programas de activación física dentro del Complejo Acuático Universitario de la Universidad Autónoma de Baja California (UABC), en Mexicali, Baja California. De este grupo, 19 personas mostraron mediante diagnóstico médico por medio de resonancia magnética, hernia de disco en la región lumbar.

### 3.3.2 Recolección de Datos.

La recolección de datos se llevó a cabo en tres pasos: primero; para cada participante se capturó información sobre su historial clínico, datos personales, antecedentes familiares, antecedentes patológicos y no patológicos así como el padecimiento y tratamiento actual; utilizando además el cuestionario Oswestry (Fairbanks et al., n.d.) y el Cuestionario Internacional de Actividad Física (IPAQ)

(CRAIG et al., 2003). Segundo; se realizó toma de glucosa (glucómetro, SD Check) y de tensión arterial (baumanómetro manual y estetoscopio, Welch Allyn). Tercero; se midió la estatura de cada individuo utilizando un estadiómetro graduado en centímetros, presentando una escala de 0 a 250 (SECA) [Medical Scales and Measuring Systems]. Finalmente, utilizando equipo InBody 720 (Biospace Co., Ltd) se realizó la captura de la composición corporal.

Con los datos recopilados se generó una base de datos de 95 variables para cada individuo; de las cuales 41 variables son de historial clínico y 54 variables son de análisis corporal. Las variables de nuestro interés son: la presencia de hernia de disco en la región lumbar y las obtenidas por el analizador corporal (InBody 720), que son: el Grado de Obesidad (GO), Área de Grasa Visceral (AGV), Índice de Masa Corporal (IMC), la Circunferencia Abdominal (CA), la Masa Muscular Esquelética (MME) y la Masa Grasa Corporal (MGC). En la Tabla 2 se observa la estadística descriptiva de las variables antes mencionadas.

**Tabla 2**. Estadística Descriptiva de las Variables GO, AGV, IMC, CA, MME y MGC.

| Variable | Media | SD | MAX | MIN | CI |
|---|---|---|---|---|---|
| Grado de Obesidad (%) | 143.62 | 28.52 | 221.69 | 92.65 | 143.4413, 143.8185 |
| Área de Grasa Visceral | 131.87 | 34.56 | 213.12 | 38.81 | 131.6433, 132.1003 |

| | | | | | |
|---|---|---|---|---|---|
| ($cm^2$) | | | | | |
| Índice de Masa Corporal ($kg/cm^2$) | 31.05 | 5.99 | 47.69 | 19.91 | 31.01303, 31.09231 |
| Circunferencia Abdominal (cm) | 104.42 | 13.68 | 140.2 | 69.8 | 104.3362, 104.5171 |
| Masa Muscular Esquelética (kg) | 24.56 | 4.45 | 38.73 | 17.13 | 24.59113, 24.53221 |
| Masa Grasa Corporal (kg) | 34.02 | 11.11 | 57.8 | 12.6 | 34.10125, 33.9543 |

SD=desviación estándar, MAX= máximo, MIN=mínimo, CI= Intervalo de confianza

### 3.3.4 Análisis de Componentes Principal (PCA)

La idea principal del Análisis de Componentes Principales (PCA) es reducir la dimensionalidad de un conjunto de datos, el cual consiste en un gran número de variables interrelacionadas, mientras retiene lo más posible de la variación presente en un conjunto de datos. Esto se logra transformando un nuevo conjunto de variables, los componentes principales (PCA), los cuales no están correlacionados, y están ordenados de forma tal que los primeros retienen la mayor variación presente en todas las variables originales (Jolliffe, 2002).

De manera formal PCA está definido como una transformación lineal ortogonal, que transforma los datos a un Nuevo Sistema de coordenadas tal, que la mayor varianza para cualquier proyección de los datos yace en la primera coordenada (llamada el primer componente principal), la segunda mayor varianza en la segunda coordenada y así sucesivamente. PCA es en teoría la transformación óptima para un conjunto de datos dado, en términos de mínimos cuadrados. El procedimiento para obtener los componentes principales puede resumirse de la siguiente manera: dado un vector $\mathbf{X^T}$ de *n* dimensiones, $X = [x_1, x_2 \ldots, x_n]^T$, del cual sus vectores de medias, **M,** y covarianzas, **C,** son descritos por: $M=E(X)= [m_1, m_2, \ldots, m_n]^T$ y $C=E [(X-M) (X-M)^T]$; calcular los eigenvalores $\lambda_1, \lambda_2, \ldots, \lambda_n$, y los eigenvectores $P_1, P_2, \ldots, P_n$; y ordenarlos de acuerdo a su magnitud $\lambda_1 \geq \lambda_2 \geq \ldots \geq \lambda_n$. Seleccionar *d* eigenvectores para representar las *n* variables, $d<n$. Entonces $P_1, P_2, \ldots, P_d$ son llamados componentes principales.

### 3.3.5 Análisis de regresión logística.

Los Modelos de Regresión son métodos matemáticos para modelar la relación estocástica cuantitativa entre una variable de interés y una (o un conjunto) de variables explicativas. De forma específica estos modelos pueden ser expresados como sigue:

$$Y_i = \beta_0 + \beta_1 X_{1i} + \beta_2 X_{2i} + \ldots + \beta_p X_{pi} + \varepsilon_i$$

donde:

$Y_i$: variable de interés, dependiente o regresando,

$X_{1i}$, $X_{2i}$,…, $X_{pi}$: variables explicativas, independientes o regresores,

$\beta_0$ : intersección o termino constante,

$\beta_1$, $\beta_2$,…, $\beta_p$: parámetros, miden la influencia que las variables explicativas tienen sobre el regresando,

*p*: número de parámetros independientes a tomar en cuenta,

$\varepsilon$ : error de observación debido a variables no controladas,

*i*: *1, 2, …, n* número de observaciones de las variables.

Con estos modelos es posible estudiar relaciones lineales entre múltiples variables y el efecto que estas tienen sobre la variable dependiente. Los $\beta_i$ se estiman siguiendo el criterio de mínimos cuadrados:

$$\min_{\substack{\beta\in n \\ i=1,\ldots,n}} \sum_{i=1}^{n} \left(Y_i - \beta_0 - \beta_1 X_{1i} - \beta_2 X_{2i} - \ldots - \beta_p X_{pi}\right)^2$$

Y los estimadores de mínimos cuadrados son obtenidos a partir de la ecuación:

$$\beta = \left(X^T X\right)^{-1} X^{-T} Y$$

En particular, la regresión logística es un tipo de análisis de regresión utilizado para predecir el resultado de una variable categórica (una variable que puede adoptar un número limitado de categorías) en función de las variables independientes o predictoras. Es útil para modelar la probabilidad de un evento ocurriendo como función de otros factores. El análisis de regresión logística se enmarca en el conjunto de Modelos Lineales Generalizados (GLM, por sus siglas en inglés) que usa como función de enlace la función *logit*. Las probabilidades que

describen el posible resultado de un único ensayo se modelan, como una función de variables explicativas, utilizando una función logística. El modelo de regresión puede expresarse como sigue(Rosner, 2011):

$$\log it(p_i) = \ln\left(\frac{p_i}{1-p_i}\right) = \beta_0 + \beta_1 X_{1i} + \beta_2 X_{2i} + ... + \beta_p X_{pi} + \varepsilon_i$$

dónde:

$$p_i = E\left(\frac{Y_i}{n_i}\middle| X_i\right)$$

### 3.3.6 Cálculo del factor de riesgo.

Los criterios para definir los factores de riesgo basados en la OR son los mencionados por Szumilas,(2010). Donde establece lo siguiente:

Si OR = 1 La variable causante no afecta el valor de la variable respuesta; no existe asociación entre las variables,

Si OR > 1 La variable causante está asociada con probabilidades altas de la variable respuesta; existe asociación entre las variables,

Si OR < 1 La variable causante está asociada con probabilidades bajas de la variable respuesta; existe asociación protectora entre las variables.

Aunado a esto, el intervalo de confidencia (CI) se utiliza para estimar la precisión del OR. Un valor de CI alto indica un bajo nivel de precisión de la OR, mientras que un valor bajo de CI indica una mayor precisión del OR. En la práctica

si el CI no contiene el valor 1, que es el valor nulo entonces decimos que la OR es significante estadísticamente (este criterio es usado como una aproximación).

### 3.4 Resultados.

#### 3.4.1 Análisis de detección de cúmulos.

El análisis de los datos se realizó con el software R versión 3.2.3 (R Development Core Team, 2021). Primero, a los 90 vectores de información generados para los individuos (ver 3.3.2), se les aplicó un corrimiento para centrarlos en cero y se normalizaron para obtener una varianza unitaria. A los vectores resultantes se les aplicó un Análisis de Componentes Principales, PCA (ver 3.3.4), con el objetivo de estudiar la estructura de los datos y ver si existen cúmulos de información en los vectores, que logren diferenciarlos. La figura 8 presenta la gráfica del Componente Principal 1 (PC1) contra el Componente Principal 2 (PC2). Los puntos de color rojo, etiquetados con un 1, corresponden a los usuarios con HDL diagnosticada, mientras que los puntos de color azul, corresponden a los usuarios sin HDL.

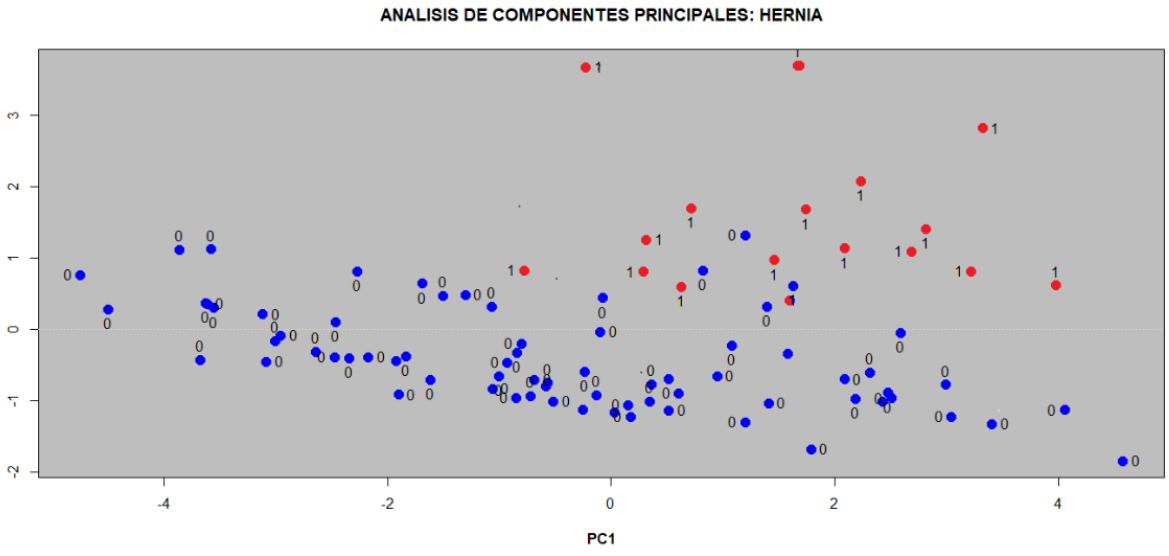

**Figura 8.** Análisis de Componentes Principales (PC1 contra PC2). Los puntos rojos corresponden a usuarios con HDL (etiquetados con un 1). Los puntos azules corresponden a usuarios sin HDL (etiquetados con un 0). Se observa diferenciación de ambos grupos de usuarios.

En la figura 8 se observa a los sujetos con HDL están concentrados en la parte superior derecha de la gráfica; con valores desde -1 y mayores para PC1, valores de 0 y mayores para PC2; mientras que los sujetos sin HDL están dispersos en todo el rango para PC1, presentando valores de 2 y menores para PC2. Podemos observar que existen algunos puntos que representan a los sujetos sin HDL que se entrelazan con los sujetos con HDL, sin embargo, se logra apreciar diferenciación entre los dos grupos.

### 3.4.2 Análisis de asociación

Con el objetivo de encontrar si existe asociación entre la HDL con las variables capturadas, se realizó un análisis de regresión logística (Ver sección de Métodos). Se utilizó como variable de interés, o regresando, la HDL, asignado un 1 a todo sujeto con el padecimiento, y un 0 a todo sujeto sin el padecimiento; y se utilizó como variables explicativas, o regresores, el GO, AGV, IMC, CA, MME y MGC. En la tabla 3 se presentan los resultados de la regresión. La columna 1 muestra la variable explicativa, la columna 2 indica la significancia estadística del efecto de la variable (*Valor p*) y la columna 3 es la clasificación debido a su nivel de significancia (* corresponde a: 0.05 > *valor p* $\geq$ 0.01; ** corresponde a: 0.01 > *valor p* $\geq$ 0.001).

**Tabla 3.** Resultados de Regresión Logística. Se muestra la variable de interés, el valor de significancia estadística de su efecto (*valor p*), y la clasificación de acuerdo a su *valor p*.

| **Variable** | **Valor p** | **Nivel de Significancia** |
|---|---|---|
| Grado de obesidad (GO) | 0.03878 | * |
| Área de Grasa Visceral (AGV) | 0.00554 | ** |
| Índice de Masa Corporal (IMC) | 0.01707 | * |
| Circunferencia Abdominal (CA) | 0.00316 | ** |
| Mas Muscular Esquelética (MME) | 0.56743 | |
| Masa Grasa Corporal (MGC) | 0.34534 | |

Código de significancia 0 “***” 0.001 “**” 0.01 “*” 0.05 “.”0.1 “” 1

A partir de los resultados de la regresión (Tabla 3), identificamos cuatro variables asociadas con la HDL. Estas son: GO, AGV, IMC y CA. Las variables GO e IMC resultaron con *valor p* de 0.03878 y 0.01707, respectivamente; mientras que las variables AGV y CA resultaron con *valor p* de: 0.00554 y 0.00316, respectivamente. Las variables MME y MGC no resultaron asociadas.

### 3.4.3 Análisis de factores de riesgo.

Para determinar si las variables asociadas son factor de riesgo, se utilizaron solo las variables que resultaron significativas en el análisis de asociación (GO, AGV, IMC y CA). Para definir factores de riesgo se aplicaron criterios basados en la Razon Odd (OR), descritos por Szumila et. al., 2010 (Szumilas, 2010). Los criterios son los siguientes:

Una variable es considerada factor de riesgo si cumple con las siguientes tres condiciones: 1) El *valor p* de la prueba estadística para la OR debe ser ≤ 0.05; 2) la OR debe ser distinta a 1; y 3) el Intervalo de Confidencia (CI) no debe incluir el valor 1. Si una variable cumple con estas tres condiciones, y su OR > 1, entonces es declarada como factor de riesgo, asociada a valores altos de la variable. Por otro lado, si la variable cumple con las tres condiciones, y su OR < 1, entonces es declarada como factor de riesgo, asociada a valores pequeños de la variable. En la tabla 4 se observa los resultados del análisis de la Razón Odds.

**Tabla 4**. Resultados del análisis de la Razón Odds para definir factores de riesgo.

| **Variable** | **Razón Odds** | **Límite inferior del Intervalo de Confidencia** | **Límite superior del Intervalo de Confidencia** | **Factor de riesgo** |
|---|---|---|---|---|
| Grado de Obesidad (GO) | 1.225870e-06 | 7.393532e-01 | 1.61866970 | No |

| | | | | |
|---|---|---|---|---|
| Área de Grasa Visceral (AGV) | 8.974036e-01 | 8.171580e-01 | 0.9683588 | Si, asociado a valores pequeños de la variable |
| Índice de Masa Corporal (IMC) | 2.636278e-01 | 6.810841e-02 | 2.8993088 | No |
| Circunferencia Abdominal (CA) | 1.415182e+00 | 1.159230e+00 | 1.8068776 | Si, asociado a valores altos de la variable |

Como se observa en la tabla 3, el AGV es un factor de riesgo para desarrollar HDL, asociada a valores pequeños de la variable; mientras que la CA es igualmente un factor de riesgo para desarrollar HDL, pero asociada a valores grandes de la variable.

## 3.5 Discusión.

Los resultados de este estudio muestran la existencia de asociación significativa del Grado de Obesidad, el Área de Grasa visceral, el Índice de Masa Corporal y la

Circunferencia Abdominal, con el desarrollo de Hernia de Disco en la Región Lumbar. El 85.22% de los pacientes de este estudio mostraron niveles de IMC con sobrepeso u obesidad siendo estos valores indicativos de un incremento de carga mecánica, lo que podría provocar como lo menciona, Iatridis et al., (2006) alteraciones sustanciales en la propiedad discal. Por su parte, Adams & Roughley, (2006)(Liuke et al., 2005) mencionan que el sobrepeso ha sido asociado con la degeneración discal a nivel histológico y macroscópico, causando cambios en la estructura del disco y afectando el proceso de regeneración, al disminuir el transporte metabólico hacia el disco.

Estudios como el reportado por Takatalo et al., (2013), quienes midieron el diámetro de adiposidad corporal mediante imágenes de resonancia magnética, además del analizador corporal InBody, encontrando la asociación entre la degeneración discal y medidas de obesidad abdominal. Estos resultados son consistentes con nuestro estudio donde el 79.44% de nuestra población, presenta rangos elevados de AGV.

En otro estudio similar, Hershkovich et al., (2013) encontraron asociación del sobrepeso y la obesidad con el dolor de espalda baja, que es un síntoma asociado a la degeneración de disco.

El Área de Grasa Visceral y la Circunferencia Abdominal resultaron como factores de riesgo para el padecimiento de HDL en la población de Mexicali, Baja California (México). Diversos estudios han mencionado como factor de riesgo la prevalencia de dolor de espalda baja (característica de la hernia de disco lumbar aunado a un dolor irradiado hacia uno de los miembros inferiores) en pacientes con IMC elevado(Chou et al., 2011), sin embargo la circunferencia abdominal y la grasa

visceral pueden jugar un rol más importante en el desencadenamiento de la HDL. Otros estudios, como el de Samartzis et al., (2012) han demostrado que el sobrepeso y la obesidad (valor evaluado por el IMC), en adultos son factores de riesgo de la presencia, extensión y severidad global de la degeneración discal. Cabe mencionar que los valores de IMC en relación a la degeneración discal, tiene numerosas limitaciones, por lo que un análisis de la composición corporal es más certero, ya que esta indica la distribución de la grasa, en particular la grasa abdominal(Samartzis et al., 2013). Takatalo et al., (2013) al realizar mediciones de adiposidad corporal mediante imágenes de resonancia magnética encontraron que los valores obtenidos eran semejantes a las mediciones de circunferencia abdominal, por lo que proponen la utilidad clínica de esta medida para evaluar la adiposidad abdominal como factor de riesgo de la degeneración discal, con la ventaja de ser un método de fácil acceso y bajo costo. Por otro lado, Han et al., (1997) mostraron que la prevalencia de dolor crónico de espalda baja y las HDL se incrementan cuando los pacientes presentan una circunferencia abdominal aumentada y un IMC elevado, sin embargo existe mayor riesgo relativo de dolor lumbar en personas con circunferencia abdominal aumentada que con los pacientes de IMC elevado.

# CAPITULO IV. CONCLUSIÓN

Este proyecto de tesis se enfocó en la aplicación de distintos métodos bioinformáticos, tanto para la identificación de variaciones estructurales basadas en LD de rango corto (VEs), como para la identificación de factores de riesgo de la composición corporal asociados a la hernia de disco lumbar, en poblaciones mexicanas, tanto nativas como mestizas. Primero, se utilizaron genotipos de alta densidad de 10 poblaciones de México (mestizas y nativas) distribuidas en las distintas zonas del territorio (centro, norte, sur y sureste); y se aplicó un método de identificación de VEs, logrando identificar 4,374. Los resultados obtenidos permitieron visualizar los distintos grupos que conformaban la población mexicana. En primera instancia, se logró generar una clara diferenciación entre poblaciones nativas y mestizas. Igualmente se logró generar una clara distinción dentro del grupo de poblaciones mestizas; tomando en cuenta su ubicación geográfica y su relación con la composición del componente ancestral.

Aunado a lo anterior, se identificó que dentro del grupo de los nativos la población Zapoteca presenta una mayor conservación de su información genética, a través del tiempo. Además, se encontraron varios genes relacionados con la salud de la población mexicana como el FTO, BRCA2, ELMO1, IGF2B2, CILP.

En segunda instancia, se realizó el análisis de asociación de la composición corporal en sujetos con hernia de disco lumbar del noroeste del país. Identificándose como asociados a la presencia de hernia de disco lumbar los siguientes cuatro componentes de la composición: el Grado de Obesidad, el Área de Grasa Visceral, el IMC y la Circunferencia Abdominal.

Sin olvidar los factores existentes, como la edad de la persona, factores mecánicos y factores genéticos, se declararon como factores de riesgo al Área de Grasa Visceral y a la Circunferencia Abdominal. Estos últimos son factores modificables, lo que permite mediante la prescripción y aplicación de programas de ejercicio físico, y de alimentación, la disminución de tejido adiposo y el incremento de la musculatura de la zona del Core con la finalidad de prevenir la aparición de la hernia de disco lumbar o como parte fundamental de su tratamiento, en la población del noroeste del país.

Este tipo de análisis remarca la importancia de conocer la conformación y estructura genética de la población, y el efecto del estilo de vida la cual puede predisponer a la aparición y desarrollo de ciertas enfermedades.

Algunas de las limitaciones presentes en el estudio fue el conocer solamente el genotipo de las 10 poblaciones, como solo conocer el fenotipo de la población del noroeste. Al contar con el fenotipo y genotipo de la población a analizar permite realizar análisis de asociación que puedan esclarecer el efecto de los genes y el estilo de vida.

Ambos análisis ofrecen un marco de referencia para futuros estudios de variaciones estructurales de poblaciones estratificadas. Igualmente, para ubicar los genes que están presentes en estas regiones variantes, y su futura asociación con enfermedades y padecimientos, tales como la Obesidad, Diabetes Mellitus y la hernia de disco lumbar, entre muchas más.

## REFERENCIAS


Abel, H. J., Larson, D. E., Regier, A. A., Chiang, C., Das, I., Kanchi, K. L., Layer, R. M., Neale, B. M., Salerno, W. J., Reeves, C., Buyske, S., Abecasis, G. R., Appelbaum, E., Baker, J., Banks, E., Bernier, R. A., Bloom, T., Boehnke, M., Boerwinkle, E., … Hall, I. M. (2020). Mapping and characterization of structural variation in 17,795 human genomes. *Nature*, *583*(7814), 83–89. https://doi.org/10.1038/s41586-020-2371-0

Adams, M. A., & Roughley, P. J. (2006). What is Intervertebral Disc Degeneration, and What Causes It? *Spine*, *31*(18), 2151–2161. https://doi.org/10.1097/01.brs.0000231761.73859.2c

Aguilar-Ordoñez, I., Pérez-Villatoro, F., García-Ortiz, H., Barajas-Olmos, F., Ballesteros-Villascán, J., González-Buenfil, R., Fresno, C., Garcíarrubio, A., Fernández-López, J. C., Tovar, H., Hernández-Lemus, E., Orozco, L., Soberón, X., & Morett, E. (2021). Whole genome variation in 27 Mexican indigenous populations, demographic and biomedical insights. *PLoS ONE*, *16*(4 April), 1–19. https://doi.org/10.1371/journal.pone.0249773

Almeida, E. P., Sabino Pinho, C. P., Leão, A. P. D., Rodrigues, I. G., Diniz, A. S., & de Arruda, I. K. G. (2018). Ratio of visceral to subcutaneous fat as a predictor of cardiometabolic risk. *Revista Chilena de Nutricion*, *45*(1), 28–36. https://doi.org/10.4067/S0717-75182018000100028

Anders Aandstad, Kristian Holthergerget, Rune Hageberg, Ingar Holme, S. A. A. (2014). Validity and Reliability of Bioelectrical Impedance Analysis and

Skinfold Thickness in Predicting Body Fat in Military Personnel. *Military Medicine*, *179*(2). https://doi.org/10.7205/MILMED-D-12-00545

Aro, S., & Leino, P. (1985). Overweight and musculoskeletal morbidity: a ten-year follow-up. *International Journal of Obesity*, *9*(4), 267–275. http://www.ncbi.nlm.nih.gov/pubmed/4066114

Avila-Arcos, C., Mcmanus, K. F., Sandoval, K., Rodríguez-Rodríguez, J. E., Villa-Islas, V., Martin, A. R., Luisi, P., Peñaloza-Espinosa, R. I., Eng, C., Huntsman, S., Burchard, E. G., Gignoux, C. R., Bustamante, C. D., & Es Moreno-Estrada, A. (2020). Population History and Gene Divergence in Native Mexicans Inferred from 76 Human Exomes. *Molecular Biology and Evolution*. https://doi.org/10.1093/molbev/msz282

Bayat, A. (2002). Clinical review Science, medicine, and the future Bioinformatics. *Bmj*, *324*(April), 1018–1022. www.ebi.ac.uk

Benjamini, Y., Drai, D., Elmer, G., Kafkafi, N., & Golani, I. (2001). Controlling the false discovery rate in behavior genetics research. *Behavioural Brain Research*, *125*(1–2), 279–284. https://doi.org/10.1016/S0166-4328(01)00297-2

Busebee, B., Ghusn, W., Cifuentes, L., & Acosta, A. (2023). Obesity: A Review of Pathophysiology and Classification. *Mayo Clinic Proceedings*, *98*(12), 1842–1857. https://doi.org/10.1016/j.mayocp.2023.05.026

Carvajal-Rodríguez, A. (2015). Incorporación de la programación informática en el currículum de Biología. *Magister*, *27*(2), 76–82.

https://doi.org/10.1016/j.magis.2015.12.002

Castaño, L., & Bilbao, J. R. (1997). Introduccion a la biologia molecular y aplicacion a la Pediatria (7): Conceptos de genetica en las enfermedades hereditarias. Genotecas. *Anales Espanoles de Pediatria*, *47*(4), 437–442.

Chama-Avilés, A., Flores-Viveros, K. L., Cabrera-Ayala, J. A., Aguilar-Galarza, A., García-Muñoz, W., Haddad-Talancón, L., de Lourdes Anzures-Cortés, M., Velázquez-Sánchez, C., Chávez-Servín, J. L., Anaya-Loyola, M. A., García-Gasca, T., Rodríguez-García, V. M., & Moreno-Celis, U. (2023). Identification and Association of Single Nucleotide Polymorphisms of the FTO Gene with Indicators of Overweight and Obesity in a Young Mexican Population. *Genes*, *14*(1). https://doi.org/10.3390/genes14010159

Chavez, Y. A., & Martinez, B. (2010). Estudios de asociación mediante rastreo genómico y su contribución en la genética del asma; Genome-wide association studies (gwas) and its contribution to the. *Salud Uninorte*, *26*(2), 269–284. http://bases.bireme.br/cgi-bin/wxislind.exe/iah/online/?IsisScript=iah/iah.xis&src=google&base=LILACS&lang=p&nextAction=lnk&exprSearch=637266&indexSearch=ID

Chou, D., Samartzis, D., Bellabarba, C., Patel, A., Luk, K. D. K., Kisser, J. M. S., & Skelly, A. C. (2011). Degenerative Magnetic Resonance Imaging Changes in Patients With Chronic Low Back Pain. *Spine*, *36*(21 Suppl), S43–S53. https://doi.org/10.1097/BRS.0b013e31822ef700

Collins, R. L., Brand, H., Karczewski, K. J., Zhao, X., Alföldi, J., Francioli, L. C.,

Khera, A. V., Lowther, C., Gauthier, L. D., Wang, H., Watts, N. A., Solomonson, M., O'Donnell-Luria, A., Baumann, A., Munshi, R., Walker, M., Whelan, C. W., Huang, Y., Brookings, T., … Talkowski, M. E. (2020). A structural variation reference for medical and population genetics. *Nature*, *581*(7809), 444–451. https://doi.org/10.1038/s41586-020-2287-8

CRAIG, C. L., MARSHALL, A. L., MICHAEL SJOSTROM, BAUMAN, A. E., BOOTH, M. L., AINSWORTH, B. E., PRATT, M., EKELUND, U., YNGVE, A., SALLIS, J. F., & OJA, P. (2003). International Physical Activity Questionnaire: 12-Country Reliability and Validity. *Medicine & Science in Sports & Exercise*, *35*(8), 1381–1395. https://doi.org/10.1249/01.MSS.0000078924.61453.FB

Eskola, P. J., Lemmelä, S., Kjaer, P., Solovieva, S., Männikkö, M., Tommerup, N., Lind-Thomsen, A., Husgafvel-Pursiainen, K., Cheung, K. M. C., Chan, D., Samartzis, D., & Karppinen, J. (2012). Genetic Association Studies in Lumbar Disc Degeneration: A Systematic Review. *PLoS ONE*, *7*(11). https://doi.org/10.1371/journal.pone.0049995

Fairbanks, J. C., Couper, J., Davies, J. B., & O 'brien, J. P. (n.d.). *Oswestry Low Back Pain Disability Questionnaire*. Retrieved January 16, 2018, from https://s3.amazonaws.com/academia.edu.documents/30399446/oswestry_low_back_pain_disability_questionnaire.pdf?AWSAccessKeyId=AKIAIWOWYYGZ2Y53UL3A&Expires=1516166619&Signature=x5h1ChOgdOseg5RDkODeL5yPHnQ%253D&response-content-disposition=inline%253B%2520filename%25

Fornoni, J., & Sosenski, P. (2009). Qué es la selección natural. *Ciencia*, *Octubre-Di*, 10–21. https://doi.org/10.1007/s11239-011-0631-4

Genetic Alliance. (2008). *Cómo entender la genética: Una guía para pacientes y profesionales médicos en la región de Nueva York y el Atlántico Medio.* https://www.ncbi.nlm.nih.gov/books/NBK132204/

González Jiménez, E. (2011). Genes y obesidad: Una relación de causa-consecuencia. *Endocrinologia y Nutricion*, *58*(9), 492–496. https://doi.org/10.1016/j.endonu.2011.06.004

González Jiménez, E. (2013). Composición corporal: Estudio y utilidad clínica. *Endocrinologia y Nutricion*, *60*(2), 69–75. https://doi.org/10.1016/j.endonu.2012.04.003

Gonzalez, S., Jiménez-López, J. C., Hedges, R., Huddart, D., Ohman, J. C., Turner, A., & Pompa y Padilla, J. A. (2003). Earliest humans in the Americas: new evidence from México. *Journal of Human Evolution*, *44*(3), 379–387. https://doi.org/10.1016/S0047-2484(03)00004-6

Guardado-estrada, M., Queipo, G., Meraz-ríos, M., & Berumen-campos, J. (2008). *Artemisa Utilización de marcadores de ADN*. *71*(3), 162–174.

Han, T. S., Schouten, J. S. A. G., Lean, M. E. J., & Seidell, J. C. (1997). The prevalence of low back pain and associations with body fatness, fat distribution and height. *International Journal of Obesity*, *21*(7), 600–607. https://doi.org/10.1038/sj.ijo.0800448

Hershkovich, O., Friedlander, A., Gordon, B., Arzi, H., Derazne, E., Tzur, D.,

Shamis, A., & Afek, A. (2013). Original Contribution Associations of Body Mass Index and Body Height With Low Back Pain in 829,791 Adolescents. *American Journal of Epidemiology*, *178*(4). https://doi.org/10.1093/aje/kwt019

Holmes, C. J., & Racette, S. B. (2021). The utility of body composition assessment in nutrition and clinical practice: an overview of current methodology. *Nutrients*, *13*(8), 1–16. https://doi.org/10.3390/nu13082493

Iatridis, J. C., MacLean, J. J., Roughley, P. J., & Alini, M. (2006). Effects of Mechanical Loading on Intervertebral Disc Metabolism In Vivo. *The Journal of Bone and Joint Surgery (American)*, *88*(suppl_2), 41. https://doi.org/10.2106/JBJS.E.01407

illumina. (2009). *SNP Genotyping and Copy Number Analysis*. /https://www.illumina.com/content/dam/illumina-marketing/documents/products/guides/2009_product_guide_snp_genotyping_cnv.pdf

illumina, I. (n.d.). *Tecnología confiable basada en perlas*. 2024. https://www.illumina.com/science/technology/microarray.html

InBody Co., L. (1996). *In Body 720 User's Manual* (pp. 1–77). InBody Bldg. https://doi.org/10.1007/978-1-4614-8687-9_1

INEGI. (2022). Estadísticas a propósito del día internacional de los pueblos indígenas. *Comunicado de Prensa*, *430(22)*(22), 1–7. https://www.inegi.org.mx/contenidos/saladeprensa/aproposito/2022/EAP_PueblosInd22.pdf

*Información genética básica - Cómo entender la genética* -. (n.d.). NCBI Bookshelf. Retrieved June 9, 2022, from https://www.ncbi.nlm.nih.gov/books/NBK132204/

Instituto Nacional de Estadística y Geografía. (2022). *Principales resultados del Censo de Población y Vivienda 2020. Estados Unidos Mexicanos*. 173.

Instituto Nacional de Pueblos Indígenas. (n.d.). *Atlas de los puebos Indígeneas de México*. 2020. https://atlas.inpi.gob.mx/mayas-etnografia/

Jeff Smith, P. D. (n.d.). *Whole-Genome Association using Mapping 500K Array Products*. https://tools.thermofisher.com/content/sfs/brochures/whole_genome_assoc_500k_jsmith.pdf

Jeffrey S. Morris1 and Veerabhadran Baladandayuthapani1. (2017). Statistical Contributions to Bioinformatics: Design, Modeling, Structure Learning, and Integration. *HHS Public Access*, *176*(1), 100–106. https://doi.org/doi:10.1177/1471082X17698255

Jiang, N., Leach, L. J., Hu, X., Potokina, E., Jia, T., Druka, A., Waugh, R., Kearsey, M. J., & Luo, Z. W. (2007). Methods for evaluating gene expression from Affymetrix microarray datasets. *BMC Bioinformatics*, *9*, 1–10. https://doi.org/10.1186/1471-2105-9-284

Jolliffe, I. T. (2002). *Principal component analysis* (2nd ed.). Springer-Verlag New York. https://doi.org/10.1007/b98835

Khan Academy. (2023). *Mecanismos de Evolución*. https://es.khanacademy.org/science/ap-biology/natural-selection/hardy-

weinberg-equilibrium/a/hardy-weinberg-mechanisms-of-evolution

Kyle, U. G., Schutz, Y., Dupertuis, Y. M., & Pichard, C. (2003). Body composition interpretation: Contributions of the fat-free mass index and the body fat mass index. *Nutrition*, *19*(7–8), 597–604. https://doi.org/10.1016/S0899-9007(03)00061-3

Le Maitre, C. L., Hoyland, J. A., & Freemont, A. J. (2007). Catabolic cytokine expression in degenerate and herniated human intervertebral discs: IL-1β and TNFα expression profile. *Arthritis Research and Therapy*, *9*(4), 1–11. https://doi.org/10.1186/ar2275

Liu, J., Chen, Y., Shan, X., & Wang, H. (2022). Association between CILP and IL-1α polymorphisms and phenotype-dependent intervertebral disc degeneration susceptibility: A meta-analysis. *Frontiers in Genetics*, *13*(October). https://doi.org/10.3389/fgene.2022.1005393

Liuke, M., Solovieva, S., Lamminen, A., Luoma, K., Leino-Arjas, P., Luukkonen, R., & Riihimäki, H. (2005). Disc degeneration of the lumbar spine in relation to overweight. *International Journal of Obesity*, *29*(8), 903–908. https://doi.org/10.1038/sj.ijo.0802974

Manelfe, C. (Claude). (1992). *Imaging of the spine and spinal cord*. Raven Press. https://books.google.com.mx/books?id=Rp99QgAACAAJ&dq=Imaging+of+the+Spine+and+Spinal+Cord&hl=es&sa=X&redir_esc=y

Martínez-Cortés, G., Salazar-Flores, J., Gabriela Fernández-Rodríguez, L., Rubi-Castellanos, R., Rodríguez-Loya, C., Velarde-Félix, J. S., Franciso Muñoz-

Valle, J., Parra-Rojas, I., & Rangel-Villalobos, H. (2012). Admixture and population structure in Mexican-Mestizos based on paternal lineages. *Journal of Human Genetics*, *57*(9), 568–574. https://doi.org/10.1038/jhg.2012.67

Mishra, P., Singh, U., & Pandey, C. M. (2019). *Application of Student ' s t - test , Analysis of Variance , and Covariance*. 407–411. https://doi.org/10.4103/aca.ACA

Moreno-Estrada, A., Gignoux, C. R., Fernández-López, J. C., Zakharia, F., Sikora, M., Contreras, A. V., Acuña-Alonzo, V., Sandoval, K., Eng, C., Romero-Hidalgo, S., Ortiz-Tello, P., Robles, V., Kenny, E. E., Nuño-Arana, I., Barquera-Lozano, R., Macín-Pérez, G., Granados-Arriola, J., Huntsman, S., Galanter, J. M., … Bustamante, C. D. (2014). The genetics of Mexico recapitulates Native American substructure and affects biomedical traits. *Science*, *344*(6189), 1280–1285. https://doi.org/10.1126/science.1251688

Moreno-Mayar, J. V., Potter, B. A., Vinner, L., Steinrücken, M., Rasmussen, S., Terhorst, J., Kamm, J. A., Albrechtsen, A., Malaspinas, A. S., Sikora, M., Reuther, J. D., Irish, J. D., Malhi, R. S., Orlando, L., Song, Y. S., Nielsen, R., Meltzer, D. J., & Willerslev, E. (2018). Terminal Pleistocene Alaskan genome reveals first founding population of Native Americans. *Nature*, *553*(7687), 203–207. https://doi.org/10.1038/nature25173

Must, A., Spadano, J., Coakley, E. H., Field, A. E., Colditz, G., & Dietz, W. H. (1999). The Disease Burden Associated With Overweight and Obesity. *JAMA*, *282*(16), 1523. https://doi.org/10.1001/jama.282.16.1523

National Human Genome Research Institute. (n.d.). *Glosario*. Retrieved October 9, 2024, from https://www.genome.gov/es/genetics-glossary

Navarrete Linares, F. (2008). *Los Pueblos Indígenas de México Pueblos Indígenas del México Contemporáneo*. http://www.cdi.gob.mx

Organizacion Mundial de la Salud. (2023). *Hipertensión*. https://www.who.int/es/news-room/fact-sheets/detail/hypertension#:~:text=Panorama general,es tomarse la tensión arterial.

Organización Mundial de la Salud. (2021). *Obesidad y sobrepeso*. https://www.who.int/es/news-room/fact-sheets/detail/obesity-and-overweight

Organización Panamericana de la Salud. (2023). *Diabetes*. https://www.paho.org/es/temas/diabetes

Paredes, J. G. (2022). Análisis de composición corporal y su uso en la práctica clínica en personas que viven con obesidad. *Revista Médica Clínica Las Condes*, *33*(6), 615–622. https://doi.org/10.1016/j.rmclc.2022.08.005

Patel, R. S., Masi, S., & Taddei, S. (2017). Understanding the role of genetics in hypertension. *European Heart Journal*, *38*(29), 2309–2312. https://doi.org/10.1093/eurheartj/ehx273

Pérez-Razo, J. C., Cano-Martínez, L. J., Vargas Alarcón, G., Canizales-Quinteros, S., Martínez-Rodríguez, N., Canto, P., Roque-Ramírez, B., Palma-Flores, C., Esteban-Martínez, R., López-Hernández, L. B., Rojano-Mejía, D., & Coral-Vázquez, R. M. (2015). Functional Polymorphism rs13306560 of the MTHFR

Gene Is Associated with Essential Hypertension in a Mexican-Mestizo Population. *Circulation: Cardiovascular Genetics*, *8*(4), 603–609. https://doi.org/10.1161/CIRCGENETICS.114.000942

Purcell, S. (2007). *PLINK*. http://pngu.mgh.harvard.edu/purcell/plink/

Python Software Foundation. (2001). *Python* (3.8.8).

R Development Core Team. (2021). *R: A Language and Environment for Statistical Computing*. R Foundation for Statistical Computing. http://www.r-project.org

Raghavan, M., Steinrücken, M., Harris, K., Schiffels, S., Rasmussen, S., DeGiorgio, M., Albrechtsen, A., Valdiosera, C., Ávila-Arcos, M. C., Malaspinas, 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). https://doi.org/10.1126/science.aab3884

Ramírez-Bello, J., Pérez-méndez, O., Ramírez-fuentes, S., Vargas-alarcón, G., & Fragoso, J. M. (2011). *Genética y genómica de la hipertensión arterial : una actualización*. *81*(3), 240–250.

Ramírez-García, S. A., Cabrera-Pivaral, C. E., Huacuja-Ruiz, L., Flores-Alvarado, L. J., Pérez-García, G., González-Rico, J. L., López-Velázquez, A., Topete-González, L. R., Rosales-Gómez, R. C., Candelario-Mejía, G., & Villa-Ruano, N. (2013). Implicaciones en la atención primaria en salud de la genética y genómica en la diabetes mellitus tipo 2. *Revista Médica Del Instituto Mexicano*

*Del Seguro Social*, *51*(3), e6–e26. http://www.redalyc.org/articulo.oa?id=457745489010

Ramirez-Garcia, S. A., Charles-Niño, C., Mazariegos-Rubí, M., Topete-González, L. R., Flores-Alvarado, L. J., Leyva Santiago, J., Mosso-González, C., Dávalos-Rodríguez, N. O., Santiago, J. L., Mosso-González, C., & Dávalos-Rodríguez, N. O. (2015). Asociación del gen ELMO1 (Snp rs1345365) con el desarrollo de diabetes mellitus tipo 2 en población mestiza Mexicana. *Invest Clini (Venezuela)*, *56*(4), 341–355. http://ve.scielo.org/pdf/ic/v56n4/art02.pdf

Reich, D., Patterson, N., Campbell, D., Tandon, A., Mazieres, S., Ray, N., Parra, M. V., Rojas, W., Duque, C., Mesa, N., García, L. F., Triana, O., Blair, S., Maestre, A., Dib, J. C., Bravi, C. M., Bailliet, G., Corach, D., Hünemeier, T., … Ruiz-Linares, A. (2012). Reconstructing Native American population history. *Nature*, *488*(7411), 370–374. https://doi.org/10.1038/nature11258

Ricci, J. A., Stewart, W. F., Chee, E., Leotta, C., Foley, K., & Hochberg, M. C. (2006). costo subsidio por diagnostico. *Spine*, *31*(26). http://journals.lww.com/spinejournal/Fulltext/2006/12150/Back_Pain_Exacerbations_and_Lost_Productive_Time.12.aspx

Romero-Hidalgo, S., Ochoa-Leyva, A., Garcíarrubio, A., Acuña-Alonzo, V., Antúnez-Argüelles, E., Balcazar-Quintero, M., Barquera-Lozano, R., Carnevale, A., Cornejo-Granados, F., Fernández-López, J. C., García-Herrera, R., García-Ortíz, H., Granados-Silvestre, Á., Granados, J., Guerrero-Romero, F., Hernández-Lemus, E., León-Mimila, P., Macín-Pérez, G.,

Martínez-Hernández, A., … Soberón, X. (2017). Demographic history and biologically relevant genetic variation of Native Mexicans inferred from whole-genome sequencing. *Nature Communications*, *8*(1). https://doi.org/10.1038/s41467-017-01194-z

Rosner, B. (2011). *Fundamentals of biostatistics* (7th ed.). Brooks/Cole, Cengage Learning.

Rubi-Castellanos, R., Martínez-Cortés, G., Muñoz-Valle, J. F., González-Martín, A., Cerda-Flores, R. M., Anaya-Palafox, M., & Rangel-Villalobos, H. (2009). Pre-hispanic Mesoamerican demography approximates the present-day ancestry of Mestizos throughout the territory of Mexico. *American Journal of Physical Anthropology*, *139*(3), 284–294. https://doi.org/10.1002/ajpa.20980

Salomon-Torres, R., Matukumalli, L. K., Van Tassell, C. P., Villa-Angulo, C., Gonzalez-Vizcarra, V. M., & Villa-Angulo, R. (2014). High density LD-based structural variations analysis in cattle genome. *PLoS ONE*, *9*(7). https://doi.org/10.1371/journal.pone.0103046

Salud Pública, I. N. (2013). *Encuesta Nacional de Salud Y Nutrición*. https://ensanut.insp.mx/encuestas/ensanut2018/informes.php

Samartzis, D., Karppinen, J., Chan, D., Luk, K. D. K., & Cheung, K. M. C. (2012). The association of lumbar intervertebral disc degeneration on magnetic resonance imaging with body mass index in overweight and obese adults: A population-based study. *Arthritis and Rheumatism*, *64*(5), 1488–1496. https://doi.org/10.1002/art.33462

Samartzis, D., Karppinen, J., Cheung, J. P. Y., & Lotz, J. (2013). Disk degeneration and low back pain: are they fat-related conditions? *Global Spine Journal*, *3*(3), 133–144. https://doi.org/10.1055/s-0033-1350054

Silva-Zolezzi, I., Hidalgo-Miranda, A., Estrada-Gil, J., Fernandez-Lopez, J. C., Uribe-Figueroa, L., Contreras, A., Balam-Ortiz, E., Del Bosque-Plata, L., Velazquez-Fernandez, D., Lara, C., Goya, R., Hernandez-Lemus, E., Davila, C., Barrientos, E., March, S., & Jimenez-Sanchez, G. (2009). Analysis of genomic diversity in Mexican Mestizo populations to develop genomic medicine in Mexico. *Proceedings of the National Academy of Sciences of the United States of America*, *106*(21), 8611–8616. https://doi.org/10.1073/pnas.0903045106

Sistema de informacion Cultural. (2023). *Tepehuanos del norte : Pueblos indígenas México.* https://sic.gob.mx/ficha.php?table=grupo_etnico&table_id=11

Slatkin, M. (2016). Linkage disequilibrium — understanding the evolutionary past and mapping the medical future. *Nature Reviews Genetics*, *9*(6), 477–485. https://doi.org/10.1038/nrg2361.Linkage

Slatkin M. (n.d.). *Flujo génico en poblaciones naturales*. 2021. https://evolution.berkeley.edu/evolibrary/article/0_0_0/evo_21_sp

Smigielski, E. M., Sirotkin, K., Ward, M., & Sherry, S. T. (2000). dbSNP: A database of single nucleotide polymorphisms. *Nucleic Acids Research*, *28*(1), 352–355. https://doi.org/10.1093/nar/28.1.352

Sudmant, P. H., Rausch, T., Gardner, E. J., Handsaker, R. E., Abyzov, A., Huddleston, J., Zhang, Y., Ye, K., Jun, G., Fritz, M. H. Y., Konkel, M. K., Malhotra, A., Stütz, A. M., Shi, X., Casale, F. P., Chen, J., Hormozdiari, F., Dayama, G., Chen, K., … Korbel, J. O. (2015). An integrated map of structural variation in 2,504 human genomes. *Nature*, *526*(7571), 75–81. https://doi.org/10.1038/nature15394

Szumilas, M. (2010). Explaining odds ratios. *Journal of the Canadian Academy of Child and Adolescent Psychiatry = Journal de l'Academie Canadienne de Psychiatrie de l'enfant et de l'adolescent*, *19*(3), 227–229. http://www.ncbi.nlm.nih.gov/pubmed/20842279

Takatalo, J., Karppinen, J., Taimela, S., Niinimäki, J., Laitinen, J., Sequeiros, R. B., Samartzis, D., Korpelainen, R., Näyhä, S., Remes, J., & Tervonen, O. (2013). Association of abdominal obesity with lumbar disc degeneration--a magnetic resonance imaging study. *PloS One*, *8*(2), e56244. https://doi.org/10.1371/journal.pone.0056244

U.S. Department of Energy Genomic Science program. (n.d.). *Human Genome Project*. https://doe-humangenomeproject.ornl.gov/wp-content/uploads/2022/08/GenomePoster2009.pdf

Universidad de Santiago de Compostela. (2024). *Unidad de Expresión Génica - Tecnología Affymetrix*. Infraestructuras de Investigación. https://www.usc.gal/es/servicios/area/infraestructuras-investigacion/plataformas/plataforma-secuenciacion-genomica-

funcional/unidad-expresion-genica-tecnologia-affymetrix

Villa-Angulo, R., Matukumalli, L. K., Gill, C. A., Choi, J., Van Tassell, C. P., & Grefenstette, J. J. (2009). High-resolution haplotype block structure in the cattle genome. *BMC Genetics*, *10*(May), 1–13. https://doi.org/10.1186/1471-2156-10-19

Villalobos-Comparán, M., Antuna-Puente, B., Villarreal-Molina, M. T., Canizales-Quinteros, S., Velázquez-Cruz, R., León-Mimila, P., Villamil-Ramírez, H., González-Barrios, J. A., Merino-García, J. L., Thompson-Bonilla, M. R., Jarquin, D., Sánchez-Hernández, O. E., Rodríguez-Arellano, M. E., Posadas-Romero, C., Vargas-Alarcón, G., Campos-Pérez, F., Quiterio, M., Salmerón-Castro, J., Carnevale, A., & Romero-Hidalgo, S. (2017). Interaction between FTO rs9939609 and the Native American-origin ABCA1 rs9282541 affects BMI in the admixed Mexican population. *BMC Medical Genetics*, *18*(1), 1–6. https://doi.org/10.1186/s12881-017-0410-y

Villegas, C., Bernal, I., Toscano, A. M., Blanquel, E., & Meyer, L. (1994). EL TIEMPO PREHISPÁNICO. In C. de México (Ed.), *Historia minima de México* (2nd ed., pp. 11–50). https://doi.org/https://doi.org/10.2307/j.ctv3f8njr.4

VisibleBody. (2024). *Under Pressure: Mechanisms of Evolution*. https://www.visiblebody.com/learn/biology/evolution/mechanisms

Woon, Y. H., Tan, P. L., & Mat Ludin, A. F. (2020). The relationship between physical activity, body mass index and body composition among students at a Pre-University Centre in Malaysia. *IIUM Medical Journal Malaysia*, *19*(2), 83–

89.

Yewen Guan, Siting Wang, Jiaqi Wang, Dihua Meng, Huihong Wu, Q. W. and H. J. (2022). Gene locus polymorphisms and expression levels of interleukin-1 in lumbar disc disease: A MOOSE-compliant meta-analysis and immunohistochemical study. *Journal of Orthopaedic Surgery and Research*, *101*(43), E31152. https://doi.org/10.1097/MD.0000000000031152

Zhu, Y., Jia, H., Li, J., Ren, S., Huang, Z., Li, F., Xing, W., Li, S., & Yang, X. (2018). Associations between Variants in BDNF/BDNFOS Gene and Lumbar Disc Herniation Risk among Han Chinese People. *Scientific Reports*, *8*(1), 6–13. https://doi.org/10.1038/s41598-018-31146-6

Zolla, C., & Márquez, E. Z. (2004). Los pueblos indígenas de México. In *Los pueblos indígenas de México*. Universidad Nacional Autónoma de México. https://doi.org/10.2307/j.ctv1xxvwr

Zonana-Nacach, A., Moreno-Cazares, M. C., & Gómez-Naranjo, R. (2013). Costo del subsidio por diagn??stico de hernia de disco en trabajadores del Instituto Mexicano del Seguro Social. *Cirugia y Cirujanos*, *81*(5). https://www.medigraphic.com/pdfs/circir/cc-2013/cc135f.pdf