

UNIVERSIDAD COMPLUTENSE DE MADRID
FACULTAD DE MEDICINA



TESIS DOCTORAL

**Relación de polimorfismos genéticos asociados a la nutrición
con el desarrollo de sobrepeso/obesidad**

MEMORIA PARA OPTAR AL GRADO DE DOCTOR

PRESENTADA POR

Cristina Martínez Santos

Directores

Antonio López Farré
José Javier Zamorano León

Madrid

© Cristina Martínez Santos, 2021

UNIVERSIDAD COMPLUTENSE DE MADRID
FACULTAD DE MEDICINA



TESIS DOCTORAL

**RELACIÓN DE POLIMORFISMOS GENÉTICOS ASOCIADOS A LA
NUTRICIÓN CON EL DESARROLLO DE SOBREPESO/OBESIDAD**

MEMORIA PARA OPTAR AL GRADO DE DOCTOR

PRESENTADA POR

CRISTINA MARTÍNEZ SANTOS

DIRECTORES

ANTONIO LÓPEZ FARRÉ Y JOSÉ JAVIER ZAMORANO LEÓN

Dedico esta tesis a mi familia

*“Investigar es ver lo que todo el mundo ha visto,
y pensar lo que nadie más ha pensado”.*

Albert Szent-Györgyi

El presente trabajo de tesis doctoral ha sido posible gracias a la implicación del Proyecto GenObIA (S2017/BMD-3773) de la Comunidad de Madrid (apoyado por fondos de la Comunidad de Madrid y por fondos los Fondos Europeos de Desarrollo Regional (FEDER) de la Unión Europea)

Agradecimientos

A mis Directores de Tesis:

El Profesor D. Antonio López Farré, de recuerdo inolvidable, promotor y artífice principal del proyecto; sin el que no habría podido llevar a cabo este proyecto. Por sus ideas, guía, dedicación y consejos gracias a los cuales he podido realizar esta Tesis. Su ausencia crea un vacío que desluce la culminación de este trabajo.

El profesor D. José Javier Zamorano León, guía cotidiano en el trabajo de la Tesis, sin cuyos consejos, correcciones, aportaciones y ayuda tampoco habría llegado a buen puerto este proyecto.

A mi Tutora de Tesis, la Profesora Paloma Ortega Molina, por su ánimo, acompañamiento y dedicación durante estos años de trabajo.

A mi esposo Manuel por su apoyo incondicional y a mi hijo Manuel, por el tiempo que les he tenido que quitar para trabajar en este estudio.

A mis padres, por acompañarme en este camino.

A la Profesora Paloma Astasio Arbiza y a todos los demás Profesores y Personal de Administración y Servicios de área de Medicina Preventiva y Salud Pública y de Secretaría de Estudiantes por su acogida, ánimos y ayuda durante estos años.

Y a todos los que de una u otra manera han hecho posible este estudio.

Muchas gracias a todos.

ÍNDICE

GLOSARIO

RESUMEN/SUMMARY.....	1
1. INTRODUCCIÓN.....	6-46
1.1.- La obesidad, la pandemia del siglo XXI.....	7
1.2.- Etiología clásica del sobrepeso/obesidad. Factores condicionantes.....	8
1.3.- Consecuencias y comorbilidades de sobrepeso y obesidad.....	16
a) Diabetes mellitus tipo 2.....	17
b) Enfermedades cardiovasculares.....	18
c) Cáncer.....	19
d) Enfermedades respiratorias.....	20
1.4.- Conceptos básicos de genética.....	21
a) La información genética (Cromatina, cromosomas y gen).....	21
b) Variabilidad genética (alteraciones genéticas).....	23
c) Nomenclatura de polimorfismos de único nucleótido (SNP).....	25
1.5.- Impacto del factor genético en el sobrepeso y obesidad.....	26
1.6.- Definición de nutrigenética. Genes relacionados	
con sobrepeso/obesidad.....	30
1.6.1.- Ciclo saciedad-hambre.....	31
a) Control hormonal del ciclo saciedad-hambre.....	32
b) Control mecánico del ciclo saciedad-hambre.....	33
c) Genética del ciclo saciedad-hambre.....	34
1.6.2.- Metabolismo de carbohidratos.....	35
1.6.3.- Metabolismo de las grasas.....	37
1.6.4.- Ejercicio físico-dieta. Genes que intervienen en la eficacia	
del ejercicio físico para la mejora de pérdida de peso.....	39
1.7.- Nuevas tecnologías de manejo de la información	
(Inteligencia artificial).....	41
1.7.1.-Big-Data.....	41
1.7.2.- Machine Learning (árboles de decisión)	43

2. HIPÓTESIS.....	47
3. OBJETIVO.....	50
4. MATERIAL Y MÉTODOS	55-62
4.1.- Población de estudio.....	57
4.2.- Medida índice de masa corporal.....	58
4.3.- Análisis genético.....	58
a) Aislamiento y cuantificación de ADN.....	58
b) Análisis de genotipado.....	59
4.4.- Generación de árboles de decisión.....	61
4.5.- Análisis estadístico.....	62
5. RESULTADOS.....	63-71
5.1.- Distribución población estudio.....	65
5.2.- Resultados de genotipado.....	66
5.3.- Aprendizaje de los árboles de decisión asociando el IMC con los polimorfismos genéticos.....	68
5.4.- Impacto de la edad y el género en el aprendizaje del árbol de decisión.....	70
6. DISCUSIÓN.....	73-83
7. CONCLUSIONES.....	85
8.- BIBLIOGRAFÍA.....	88-106
9.- ANEXO I.....	108

GLOSARIO

ADN: Ácido desoxirribonucleico

ADBR2: Receptor adrenérgico beta 2 (beta2-adrenergic receptor)

ADBR3: Receptor adrenérgico beta 3 (beta3-adrenergic receptor)

AdipoQ: Adiponectina (Adiponectin)

AGRP: Proteína r-agouti (Agouti-related protein)

ATP: Trifosfato de adenosina (Adenosine triphosphate)

FTO: Alpha-ketoglutarate-dependent dioxygenase FTO (FTO Dioxigenasa dependiente de alfa-cetoglutarato)

GWAS: Estudio de asociación del genoma completo (Genome Wide Association Study)

HWE: Equilibrio Hardy-Weinberg.

IMC: Índice de Masa Corporal

INSIG2: Gen 2 inducido por insulina (insulin induced gene 2)

LEP: Leptina (Leptin)

LEPR: Receptor de leptina (Leptin receptor)

MC4R: Receptor de melanocortina 4 (melanocortin 4 receptor)

OECD: Organización para el Desarrollo y la Cooperación Económica (Organization for Economic Cooperation and Development)

PPARG: Receptor gamma activado por proliferador de peroxisomas (Peroxisome proliferator-activated receptor gamma)

SNP: Polimorfismo de único nucleótido (Single Nucleotide Polymorphism)

UCP2: Proteína desacoplamiento mitocondrial 2 (uncoupling protein 2)

RESUMEN

Introducción: La obesidad es una de las enfermedades más prevalentes en países industrialmente desarrollados, considerándose una de las cuatro pandemias del tercer milenio. El sobrepeso/obesidad es causa y consecuencia de otras patologías derivadas como las enfermedades cardiovasculares y metabólicas, diabetes, cáncer, etc. Por ello, la prevención y la predicción de la obesidad se ha convertido en uno de los principales objetivos de la comunidad médico-científica en las últimas décadas. En la actualidad se conoce el papel crucial que juega el aspecto genético en la obesidad. Sin embargo, el control genético de las formas comunes de obesidad en el humano no se comprende bien. La genética, y en concreto la nutrigenética, nos permite valorar el efecto de diferentes variantes genéticas sobre aspectos íntimamente ligados con el riesgo de obesidad como el control del ciclo saciedad-hambre, metabolización de hidratos de carbono, grasas, etc. y, por tanto, establecer estrategias para prevenir y disminuir el exceso de peso corporal. El imparable desarrollo de las nuevas tecnologías de análisis de datos por inteligencia artificial, como los árboles de decisión basados en el aprendizaje automático, aportan nuevas herramientas para la predicción de enfermedades. En este sentido, los árboles de decisión, a través del análisis de forma conjunta de las variables no modificables como edad, género y variantes genéticas, podrían estratificar y predecir el riesgo de sobrepeso/obesidad, con el objetivo de diseñar recomendaciones para la prevención de sobrepeso y obesidad.

Objetivos: El objetivo del trabajo doctoral fue desarrollar un análisis piloto basado en la generación de árboles de decisión para predecir el efecto de las interacciones de género, edad y diferentes variantes genéticas (single nucleotide polymorphisms, SNP) en genes implicados en diferentes aspectos nutricionales como control ciclo saciedad/hambre, metabolismo de carbohidratos y grasas y consumo calórico durante el ejercicio físico para predecir el riesgo de desarrollar sobrepeso/obesidad.

Materiales y métodos: Se incluyeron 151 individuos, seleccionados de forma aleatoria de una cohorte anónima de 1500 individuos pertenecientes a población rural demográficamente estable de Castilla La Mancha. En la población se recogieron datos de edad, género y se midió el índice de masa corporal (IMC). A partir de la obtención de una muestra de ADN de la mucosa de la cavidad oral, se analizaron diferentes polimorfismos de genes asociados al ciclo saciedad/hambre, al metabolismo de nutrientes y al gasto de calorías por ejercicio físico, mediante el método de genotipado por PCR cuantitativa a tiempo real utilizando sondas Taqman. La determinación de equilibrio de distribución de la frecuencia alélica de las variantes analizada se estimó mediante el equilibrio Hardy-Weinberg (HWE), basado en un test chi cuadrado. El resto de variables se sometió un test de normalidad Kolmogorov. Las variables continuas (edad) se compararon entre los grupos de interés mediante el test paramétrico de T de Student. Los datos de variables continuas se presentan como la media \pm EEM, a no ser que se indique de otra forma. Las diferencias en la distribución de los diferentes

genotipos se estimaron mediante el test de chi-cuadrado, ajustando por Fisher en los casos necesarios. Las diferencias se consideraron estadísticamente significativas cuando el valor $P < 0,05$. El procesado de datos y el análisis estadístico se llevó a cabo con el programa estadístico SPSS (versión 25.0). Adicionalmente a los análisis estadísticos, se generaron árboles de decisión con las variables edad, género, variantes genéticas e IMC, mediante el algoritmo C4.5 y representados en forma de gráficos.

Resultados: Los resultados revelaron que la información obtenida del análisis estadístico convencional de la distribución de sobrepeso/obesidad y las variantes genéticas analizadas era menor que la derivada del aprendizaje con árboles de decisión. La variante alélica en homocigosis Ala55Ala del gen UCP2 podría ser interpretada como un protector frente a sobrepeso/obesidad en la población analizada en este estudio, mientras que las variantes Val55Val y Val55Ala incrementaron el riesgo de sobrepeso/obesidad. El aprendizaje del árbol de decisión identificó un haplotipo específico asociado con alto riesgo de sobrepeso / obesidad. Los portadores de las variantes alélicas AgRP (rs5030980) Ala67Ala, ADRB2 (rs1042714) Gln27Glu o Glu27Glu, INSIG2 (rs7566605) 73+9802 con genotipos CC o GG y PPARG (rs1801282) con las variaciones de alelos de Ala12Ala o Pro12Pro desarrollaron con mayor probabilidad sobrepeso/obesidad $IMC \geq 25 \text{ Kg/m}^2$, respecto a los no portadores de las variantes anteriores. Por otro lado, cuando se analizaron los efectos de la interacción del género, edad y variantes genéticas sobre el riesgo de sobrepeso/obesidad, los resultados revelaron que los árboles de decisión permitieron predecir el riesgo de desarrollar sobrepeso/obesidad, estratificando por las características genéticas, género y edad. En el género femenino, la presencia de las variantes alélicas CT en el polimorfismo AgRP rs5030980 y GG en el polimorfismo genético INSIG2 rs7566605 y una edad no incluidas en los cortes etarios $>38,5$ y $\geq 70,5$ años, tenían mayor riesgo de sobrepeso/obesidad que las mujeres no portadoras de las anteriores variantes descritos y edades inferiores a 38,5 años.

Conclusiones: La generación de árboles de decisión (técnica de machine learning) permite analizar de forma conjunta variables no modificables como el género, edad y variantes genéticas de alta frecuencia en la población (polimorfismo), identificando patrones de predicción de riesgo de sobrepeso/obesidad y, por tanto, la aparición de comorbilidades asociadas. Este estudio respalda el amplio campo de aplicación de las nuevas tecnologías de análisis de datos en el campo de la medicina predictiva.

Palabras clave: Árboles de decisión, Machine-learning, inteligencia artificial, polimorfismos genéticos, nutrigenética, índice de masa corporal, normopeso, sobrepeso, obesidad.

SUMMARY

Introduction: Obesity is one of the most prevalent diseases in developed countries, being considered one of the four main pandemics of the third millennium. Overweight/obesity is both cause and consequence of other subsequent conditions such as cardiovascular and metabolic diseases, diabetes, cancer amongst others. Therefore, predicting and preventing obesity is one of the main goals for the scientific community in recent decades. Nowadays the crucial role that genetics play in obesity is well known. However, how genetics control and influence common obesity is not fully understood. Genetics, and more specific nutrigenetics, enables us to assess the effect of different genetic variants on aspects that are directly linked with obesity such as satiety-hunger cycle control, fat, metabolism of carbohydrates, etc. And therefore, set different strategies to prevent and reduce excess body weight. The unstoppable development of new technologies for data analysis by artificial intelligence, such as decision trees based on automatic learning, provide new tools for the prediction of diseases. In this sense, decision trees, through joint analysis of non-modifiable variables such as age, gender and genetic variants, could stratify and predict the risk of overweight/obesity, with the aim of designing recommendations for the prevention of overweight and obesity.

Objectives: The main aim of this doctoral project was to develop a pilot analysis based on decision trees generation to predict the effect of gender, age and different genetics variants (single nucleotide polymorphism, SNP) interactions in genes involved in different nutritional aspects such as satiety-hunger cycle control, metabolism of fat and carbohydrates and caloric consumption during exercise to predict the risk to develop overweight/obesity.

Materials and methods: 151 subjects were randomly selected from a cohort of 1500 anonymous individuals cohort from a rural population demographically stable in Castilla La Mancha. Age and gender were collected and body mass index (BMI) was measured. A DNA sample was collected from the mucosa of the oral cavity, and different gene polymorphisms linked to the satiety-hunger cycle, nutrients metabolism and calorie expenditure were analyzed using the genotyping method with real-time quantitative PCR, using Taqman probes. The balance of the distribution of allelic frequency of the analyzed variants was estimated through the Hardy-Weinberg (HWE) balance, based on a chi-squared test. The other variables were analysed using the Kolmogorov normality test. T parametric test was used between interest groups to compare continuous variables (age). Continuous variable data are shown as mean \pm MSE unless another method is indicated. A Chi-square test was used to identify differences in the distribution of the different genotypes, using the Fisher post hoc test when necessary. Differences were considered statistically significant when $P < 0.05$. SPSS software package (v. 25.0) was used for the data process and statistical analysis. Further to the statistical analysis,

decision trees were generated with age, gender, genetic variants and BMI variables using the C4.5 algorithm and displayed in graphics.

Results: Results showed that the information obtained from the conventional statistic analysis of the distribution of overweight/obesity and the genetic variants analyzed was less than that derived from learning with decision trees. The allelic variant in the homozygous Ala55Ala of the UCP2 gene could be interpreted as a protector against overweight/obesity in the population analyzed in this study, while the Val55Val and Val55Ala variants increased the risk of overweight/obesity. The decision tree learning identified a specific haplotype linked to a high risk of overweight/obesity. Carriers of allelic variants AgRP (rs5030980) Ala67Ala, ADRB2 (rs1042714) Gln27Glu o Glu27Glu, INSIG2 (rs7566605) 73+9802 with CC o GG y PPARG (rs1801282) genotypes with the allelic variations Ala12Ala or Pro12Pro had a higher probability to develop overweight/obesity BMI \geq 25 Kg/m² than non-carriers. On the other hand, when effects of the gender, age and genetic variants interaction were analyzed for the risk of overweight/obesity, results revealed that decision trees allowed to predict the risk of developing overweight/obesity, stratifying by genetic features, gender and age. Female gender, in presence of allelic variants CT in the AgRP rs5030980 polymorphism and GG in the INSIG2 rs7566605 polymorphism, and with an age out of age cuts >38.5 and \geq 70.5 years old, had more overweight/obesity risk than women non-carriers of previous variants and younger than 38.5 years old.

Conclusion: The generation of decision trees (machine learning technique) makes it possible to jointly analyze non-modifiable variables such as gender, age and high-frequency genetic variants in the population (polymorphism), identifying prediction patterns of risk of overweight/obesity and, therefore, the appearance of associated comorbidities. This study supports the broad field of application of new data analysis technologies in the field of predictive medicine.

Keywords: Decision trees, Machine-learning, artificial intelligence, genetic polymorphisms, nutrigenetics, body mass index, normoweight, overweight, obesity.

1. INTRODUCCIÓN

1.1.- LA OBESIDAD, LA PANDEMIA DEL SIGLO XXI

El sobrepeso y la obesidad es un grave problema de salud pública, principalmente en países industrialmente desarrollados, que se asocia con efectos adversos muy graves para la salud de población. En la actualidad, se conocen numerosos estudios que han demostrado la asociación de obesidad con diferentes patologías crónicas graves, como enfermedades cardiovasculares, respiratorias e incluso oncológicas (1-3). Curiosamente, uno de los estudios más llamativos acerca de la estrecha relación entre obesidad y el riesgo de mortalidad fue realizado a mediados del siglo XX, en el que las aseguradoras elaboraron unas tablas de riesgos específicos de padecer enfermedad cardiovascular y muerte según el peso de los sujetos. No obstante, esta aproximación de riesgo no se realizó por profesionales sanitarios, si no por agentes de compañías de seguros, con el propósito de optimizar las primas de los asegurados acorde a su riesgo específico de mortalidad (4).

El índice de masa corporal (IMC) es uno de los parámetros más utilizados para valorar el sobrepeso y la obesidad. Este índice se calcula dividiendo la masa corporal obtenida en kilogramos entre el resultado de elevar al cuadrado la altura del individuo expresada en metros (kg/m^2). Así, un individuo que presente un valor del IMC comprendido entre 18,5 y 24,9 kg/m^2 , se considera normopeso, mientras que para valores del IMC entre 25 y 30 kg/m^2 se define como sobrepeso y para valores superiores a 30 kg/m^2 como obesidad (5).

Según la Organización Mundial de la Salud (OMS), el número de personas obesas se ha triplicado desde 1975, estimándose que en 2016 había aproximadamente 1,9 billones de adultos con sobrepeso a nivel mundial, lo que supone casi un 30% de la población mundial; de los cuales 650 millones eran obesos (aproximadamente un 5% de la población mundial). Diferentes trabajos han estimado que, si la incidencia se mantiene, en 2030 más de la mitad de la población mundial tendrá un peso superior al recomendado (un 38% de la población con sobrepeso y 20% con obesidad) (6-8). El sobrepeso en Europa afecta entre un 30% y un 80% de la población adulta y los costes directos de los tratamientos de enfermedades derivadas de la obesidad, como pueden ser la diabetes tipo 2, ciertos tipos de cánceres y otras enfermedades cardiovasculares,

ascienden aproximadamente a un 6% del gasto total en salud (9). Además, también habría que tener en cuenta los costes indirectos debidos a baja productividad, absentismo, muerte prematura y los costes intangibles, que se derivan de la pérdida de oportunidades, problemas psicológicos y una peor calidad de vida (9).

En España, según la Organización para la Cooperación y el Desarrollo Económico (OECD), se estimó que en 2015 el 16,7% de la población mayor de 15 años era obesa, siendo algo superior el porcentaje de mujeres obesas que el de hombres. Actualmente, se ha estimado que el porcentaje con obesidad alcanzará el 21% de la población mayor de 15 años en el año 2030 (10). Un estudio en el que se realizó una estimación de la evolución del exceso de peso a partir de los datos obtenidos entre 1987 y 2014 en España, reveló que el IMC aumentó 0.1 puntos/año en hombres y 0.26 en mujeres/año (11). Este “pequeño aumento”, supuso un sobrecoste médico directo de 1950 millones de euros al año (un 2% del presupuesto sanitario) (11). De forma preocupante, si se mantuviera esta tendencia creciente hasta 2030, el sobrecoste médico directo aumentaría un 58% (unos 3.000 millones de euros anuales), totalmente inasumible para el sistema nacional sanitario, y el número de casos de exceso de peso aumentaría un 16% (unos 3,1 millones de nuevos casos) (11).

Con los datos anteriores, la obesidad se ha definido como la pandemia del siglo XXI (12), poniendo de manifiesto la imperiosa necesidad de elaborar acciones conjuntas a nivel político y médico-científico contra la obesidad, a través del desarrollo de diferentes programas de formación y concienciación, así como la implantación de intervenciones alimentarias y nutricionales específicas y personalizadas, basadas en nuevas estrategias de investigación que aporten conocimientos sobre la etiología de la obesidad y sus comorbilidades.

1.2.- ETIOLOGÍA CLÁSICA DEL SOBREPESO/OBESIDAD. FACTORES CONDICIONANTES

El sobrepeso y la obesidad son el resultado de un desequilibrio energético. Considerando que para estar en equilibrio la energía que se obtiene de los alimentos debe ser igual a la energía consumida en la actividad física y en el metabolismo. Por tanto, si se ingiere una mayor cantidad de energía de la que se consume, ésta será

almacenada en forma de tejido graso en el organismo, resultando en la aparición de sobrepeso u obesidad (5,13).

Aunque esta relación a priori puede parecer sencilla, realmente existen múltiples factores que afectan o condicionan este balance, entre ellos podemos destacar:

- **Hábitos alimentarios**

El tipo y las características de los alimentos consumidos van a condicionar la energía aportada y, por tanto, el riesgo de sobrepeso y obesidad. Por ejemplo, el consumo de bebidas azucaradas, comida rápida o “*chatarra*”, carnes rojas, derivados cárnicos altamente procesados, bollería industrial, etc., van a incrementar en gran medida el riesgo de sobrepeso y obesidad. Por el contrario, un patrón de alimentación basado en el consumo de legumbres, cereales integrales con alto contenido en fibra que favorece el efecto de saciedad y disminuye el índice glucémico, verduras, frutas y frutos secos como nueces o almendras, entre otros alimentos, disminuyen el riesgo de sobrepeso y obesidad (13). Recientemente, también se ha descrito que la interacción entre los nutrientes y los genes durante el período de gestación condicionará la capacidad que tendrá un individuo en su período adulto para producir tejido adiposo y, por tanto, el riesgo de desarrollar sobrepeso/obesidad (9).

- **Factores socioeconómicos**

Diferentes factores socioeconómicos como el nivel de formación educativa, la clase social o el poder adquisitivo también se encuentran íntimamente relacionados con el riesgo de obesidad (14). Según la OECD, los hombres con un nivel educativo mayor tienden a tener menos problemas de obesidad, mientras que la diferencia es mucho mayor en las mujeres; en España en el año 2014 había casi cuatro veces más mujeres obesas con un bajo nivel educativo que con un alto nivel educativo (10).

El nivel económico de la población también influye, la calidad de los alimentos consumidos depende en gran medida del poder adquisitivo para poder tener acceso a alimentos frescos y saludables y poder evitar la comida rápida que tiene un coste mucho menor que las saludables (9).

- **Factores culturales y sociales**

La sociedad en la que vivimos, enmarcada en un entorno socio-cultural definido, va a determinar en gran medida el comportamiento de una persona, sus hábitos alimentarios, estilo de vida y finalmente el riesgo de sobrepeso u obesidad (5). Son muchos los factores socio-culturales que pueden modular el riesgo de sobrepeso/obesidad. Entre ellos pueden destacar el lugar de residencia, los residentes de regiones sometidas a baja temperatura durante prolongados períodos de tiempo presentan un aumento de la tasa metabólica basal para mantenerse caliente (aumenta el consumo de energía y, por tanto, de calorías), sin embargo, hoy en día, con la climatización en los hogares y lugares de trabajo este gasto de calorías de adaptación al medio se ha reducido (15).

En la sociedad actual, el elevado número de horas dedicadas al trabajo favorece la compra de productos procesados poco saludables, perdiéndose la saludable dieta mediterránea por otros patrones alimentarios que conllevan menor tiempo en la preparación de alimentos, pero que son claramente menos saludables y obesigénicos (9,10). Sorprendentemente, el número de horas de sueño también afecta a los valores del IMC de forma inversamente proporcional, cuando hay falta de sueño crónica existe un aumento del riesgo de obesidad (15). Los desórdenes del sueño también se han asociado a desequilibrios hormonales y cambios metabólicos que favorecen el apetito (16).

- **Tabaquismo**

El tabaco influye en la obesidad, los no fumadores suelen tener un IMC mayor que los fumadores, pero al dejar de fumar se tiende a engordar. Esto se debe a que fumar disminuye el apetito y aumenta la tasa de metabolismo basal. Se estima que fumar 24 cigarrillos al día produce un aumento del metabolismo basal de unas 200 kcal/día (5,15).

- **Factores psicológicos**

Factores psicológicos como la depresión, la desmotivación, el estrés, etc. favorecen la aparición de trastornos alimentarios como el atracón, e incluso, una pérdida de la percepción de la propia figura que favorecen el sobrepeso y la obesidad (14). Asimismo, los tratamientos farmacológicos asociados a los trastornos anteriormente señalados

como antidepresivos y antipsicóticos favorecen por sí mismos el sobrepeso y la obesidad (9,15).

- **Ejercicio físico**

La realización de ejercicio físico supone un consumo energético a tener en cuenta en el balance entre la energía gastada y la energía aportada por la alimentación, por ello, llevar una vida sedentaria supone un factor de riesgo y favorece el aumento de peso al disminuir el gasto energético. Realizar actividad física de forma regular aumenta el consumo energético favoreciendo la pérdida de peso o su equilibrio (5,13-16). Existen numerosos estudios en los que se demuestra que la combinación de la actividad física y una dieta saludable favorece la pérdida de peso (5,13-16). Sin embargo, si se quiere perder peso exclusivamente mediante el ejercicio físico, hay numerosos estudios que coinciden en que la pérdida de peso es escasa, en torno a unos 3 kg, dependiendo de la intensidad y duración del ejercicio realizado (17,18).

Cada individuo responde de diferente forma ante el ejercicio físico (19). Así, una persona al realizar ejercicio físico durante un tiempo determinado puede perder peso, mientras que otro individuo realizando la misma actividad puede mantenerse en el mismo peso (20). El ejercicio físico estimula respuestas en el tejido adiposo y en el muscular, pero las respuestas en intensidad no son iguales en todas las personas. Estas respuestas diferenciales de cada individuo al ejercicio físico se deben al papel del componente genético (16). La presencia de alteraciones en la secuencia de diferentes genes, como ADBR2 y ADBR3, juegan un papel crucial en la cantidad de energía gastada durante el ejercicio físico (21,22). Posteriormente, en un apartado específico se profundizará más en la influencia de estos genes y el ejercicio físico en el control del peso.

- **Edad**

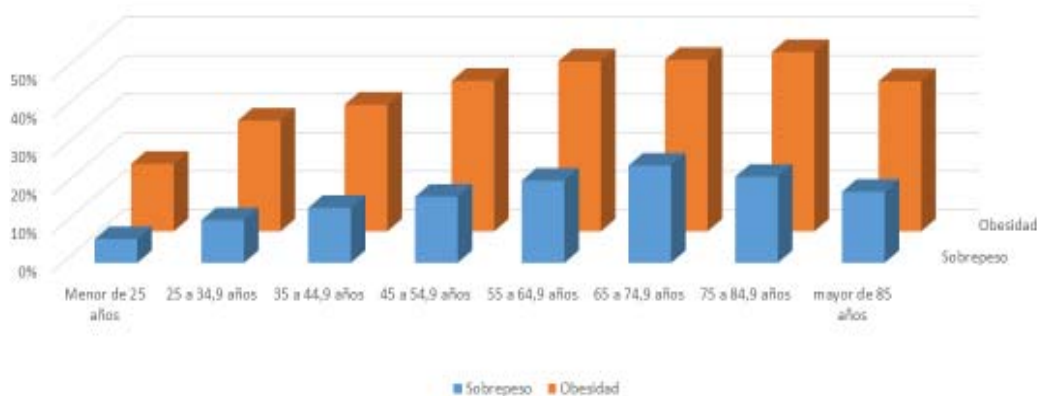
La edad influye en la predisposición a sobrepeso/obesidad, así las etapas críticas en las que se puede producir un aumento de peso son:

- *Prenatal*: Existen evidencias que sugieren que el desarrollo en el útero tiene efectos permanentes sobre el crecimiento posterior y la regulación de la energía.

- *5-7 años*: los hábitos de alimentación cambian debido a la incorporación a un entorno escolar y la interacción con otros niños.
- *Adolescencia*: al tener mayor autonomía, se producen cambios en los hábitos alimenticios con comidas más irregulares y la percepción de la figura, especialmente en mujeres.
- *Juventud*: en esta etapa se tiende a reducir el ejercicio físico significativamente, en mujeres suele ocurrir entre los 15-19 años, y en los hombres suele ser más tarde, llegando incluso hasta casi los 30 años.
- *Embarazo*: el exceso de peso adquirido durante el embarazo, suele mantenerse después del parto.
- *Menopausia*: durante la menopausia las mujeres tienden a ganar peso, y eso se suma a que la tasa del metabolismo se reduce ligeramente
- *Envejecimiento*: aumenta la grasa visceral sin un aparente cambio en el IMC, la altura disminuye afectando así al valor del IMC, aumenta el tejido graso y disminuye el muscular.

En la figura 1 se muestran los resultados de la distribución de obesidad según edad y género recogidos en la Encuesta Europea de Salud en España 2017, en la que se puede apreciar claramente la evolución progresiva de sobrepeso y obesidad según el aumento de edad hasta los 65 años, momento en el cual la proporción de obesos se estabiliza y comienza a disminuir (23). En el caso del sobrepeso se ve que aumenta hasta los 75 años, momento en el cual pasa a estabilizarse y después comienza a descender (23).

Figura 1. Histograma de sobrepeso y obesidad en función de la edad (datos obtenidos de la Encuesta Europea de Salud en España 2017).



Una de las causas de los cambios en el peso corporal según aumenta la edad son los niveles de hormonas sexuales (andrógenos y estrógenos), que varían en las diferentes etapas de la vida. En la pubertad aumentan tanto en el caso de los hombres como en el de las mujeres. Cuando las mujeres se encuentran en edad fértil, estos niveles varían mensualmente según las diferentes etapas del ciclo menstrual, al igual que durante el embarazo y la lactancia, etapas en las cuales es necesario un mayor aporte energético. Tanto para los hombres como para las mujeres, los niveles de hormonas sexuales disminuyen progresivamente en edades avanzadas. En el caso de los hombres, la testosterona disminuye a partir de los 40 años aproximadamente un 2% al año, mientras que el 95% de las mujeres tienen la menopausia entre los 45-55 años, momento en el cual, se produce un gran cambio hormonal al desaparecer el ciclo menstrual (24-26). Como se describirá en el siguiente apartado, las hormonas sexuales tienen una importante función reguladora del balance energético en el cuerpo, por lo que estas variaciones afectan, favoreciendo la acumulación de grasa visceral según disminuye la cantidad de hormonas producidas.

- **Género**

Los hombres tienen mayor tendencia a la obesidad y al sobrepeso que las mujeres hasta los 65 años aproximadamente, momento en el que las mujeres pasan a superar el porcentaje de los hombres (27). En España, los datos de la Encuesta Europea de Salud en España 2017 revelaron que la prevalencia de sobrepeso y obesidad en hombres era significativamente superior al de las mujeres (23). De estos datos se desprende que en hombres hay un 45,5% que sufren sobrepeso, frente a un 31,1% de mujeres con sobrepeso. Respecto a la obesidad, la prevalencia fue de 18,5% en hombres frente a un 17,1% en mujeres (23) (Tabla 1). En cambio, la proporción de hombres y de mujeres dentro de la población con obesidad, el 48,8% son hombres y el 51,2% son mujeres, invirtiéndose el resultado (23) (Tabla 1).

Tabla 1. Clasificación de sobrepeso y obesidad en función del sexo y del IMC (datos de la Encuesta Europea de Salud en España 2017).

		(IMC) del adulto		
		Normopeso	Sobrepeso	Obesidad
Hombre	% dentro del género	35,1%	45,5%	18,5%
	% dentro del IMC del adulto	39,0%	56,3%	48,8%
Mujer	% dentro del género	48,5%	31,1%	17,1%
	% dentro del IMC del adulto	61,0%	43,7%	51,2%
TOTAL		42,2%	37,8%	17,8%

Los cambios hormonales que sufren las mujeres durante el ciclo menstrual, el embarazo y la menopausia, afectan al equilibrio energético del cuerpo. El porcentaje de hormonas sexuales como la testosterona, la progesterona y los estrógenos es diferente en los hombres y en las mujeres, lo que produce diferencias tanto en el consumo como en el gasto energético (25,28,29). Los hombres tienen 20 veces más testosterona que las mujeres, mientras que las mujeres en edad fértil tienen unos 200 veces más estrógenos que los hombres. La función de los estrógenos en los hombres y las mujeres también es diferente, ya que el estradiol en los hombres se sintetiza fuera de las gónadas y tiene una función paracrina y autocrina, mientras que en las mujeres se sintetiza en los ovarios y funciona como una hormona circulante hasta la menopausia, momento en el cual pasa sintetizarse y funcionar igual que en los hombres. Las hormonas sexuales en general se unen menos a la globulina en los hombres que en las mujeres por lo que la cantidad de hormonas libres en sangre también difieren entre el género (30).

Las hormonas sexuales, también regulan la homeostasis energética y procesos neuronales en el sistema nervioso central. En el sistema nervioso periférico, estas hormonas, y en particular los estrógenos, tienen potenciales efectos protectores en el metabolismo de la glucosa y los lípidos en el hígado y juegan un papel crítico en el desarrollo de algunos tipos de cáncer (31). En la siguiente tabla 2, basada en los

resultados del trabajo de Alessandra Gambineri y Carla Pelusi (30), se describe cómo las hormonas masculinas y las femeninas tienen diferentes funciones según el tejido sobre el que actúan.

Tabla 2. Efectos de hormonas andrógenas y estrógenas sobre el sobrepeso/obesidad. (Tabla traducida de Alessandra Gambineri y Carla Pelusi (30).

EFECTO DE ANDRÓGENOS	LUGAR DE ACTUACIÓN	EFECTO DE ESTRÓGENOS
	Hipotálamo	↓ Ingesta de alimentos ↑ Gasto energético Regula la distribución de grasa corporal
↑ Hipertrofia de los adipocitos ↓ Sensibilidad a la insulina ↓ Almacenamiento de lípidos ↑ Lipolisis	Tejido adiposo	↓ Acumulación de tejido adiposo blanco ↓ Lipolisis
	Células β	Protege estas células frente a daños ↑ Síntesis de insulina estimulada por glucosa
	Hígado	Mantiene la acción de la insulina
↑ Síntesis de proteínas ↑ Diferenciación de células madre en miotubos ↑ Sensibilidad a la insulina ↑ Uso de glucosa ↑ Oxidación de lípidos	Músculo esquelético	↓ Consumo de ácidos grasos ↑ Oxidación de ácidos grasos Mantiene la acción de la insulina
	Macrófagos	↓ Inflamación del tejido ↑ Sensibilidad a la insulina periférica

A parte de las evidentes diferencias fisiológicas entre hombres y mujeres, las conductas son diferentes, los hombres suelen consumir más alcohol, las mujeres suelen comer más dulces, mientras que los hombres consumen más alimentos ricos en grasas de origen animal. En cambio, el ejercicio físico suele ser mayor por parte de los hombres

que por parte de las mujeres, que tienden a llevar una vida más sedentaria. Esto se debe en parte a que los trabajos que conllevan esfuerzo físico suelen realizarse principalmente por hombres, y fuera del trabajo, aunque hay individuos de los dos géneros que realizan actividades físicas, la proporción de hombres parece ser también mayor (28,29).

- **Factores genéticos**

La genética tiene una gran influencia en el desarrollo del sobrepeso y la obesidad, hoy en día se sabe que existen más de cien genes que tienen relación con la ganancia de peso, influyendo así directamente en el número de individuos que sufren o pueden sufrir sobrepeso y obesidad. De hecho, es ampliamente conocida la transmisión hereditaria de la obesidad, sin embargo, hay muchos componentes genéticos que aún no han sido estudiados (32).

Estudios basados en gemelos homocigotos han revelado que independientemente de que hayan sido criados juntos o no, tienen un IMC muy similar (33). En el caso de los gemelos heterocigotos, su IMC también guarda relación, aunque en menor medida que para los gemelos homocigotos. Esta relación es independiente de los factores ambientales que también influyen sobre el sobrepeso y la obesidad, a pesar de que tienen relación tanto con el consumo como con el gasto energético influyendo directamente sobre el IMC como se ha comentado anteriormente (33). Otro de los aspectos que evidencia la importancia del factor genético sobre el control del peso es la respuesta diferencial en el peso y composición corporal de masa magra y grasa de cada individuo a idéntica dieta (34).

Existen numerosos genes que afectan al desarrollo del sobrepeso o/y obesidad, destacan los genes FTO, LEP, LEPR, MC4R, UCP2, AgRP, PPARG, ADRB3, ADRB2 e INSIG2 entre otros, que están implicadas en diferentes aspectos relacionados con el ciclo apetito/saciedad, metabolización de carbohidratos o grasas, termorregulación, etc (35).

1.3.- CONSECUENCIAS Y COMORBILIDADES DE SOBREPESO Y OBESIDAD

Independientemente del género, la raza o la edad, el sobrepeso y la obesidad se asocia significativamente con un aumento del riesgo de mortalidad. Según el estudio de Kenneth y colaboradores, el riesgo de mortalidad en personas que sufren sobrepeso a

los 50 años aumenta entre un 20-40% respecto a individuos con normopeso a la misma edad (36).

Existen diferentes estudios poblacionales que confirman el hecho de la estrecha relación entre mortalidad y sobrepeso/obesidad (37). Por ejemplo, según Banegas et al., en Europa una de cada trece muertes era atribuible a la obesidad. De entre esas muertes, aproximadamente un 70% de ellas se debían a problemas cardiovasculares y un 20% a enfermedades oncológicas (38). Es por este motivo, por lo que la obesidad se considera el séptimo factor de riesgo que más mortalidad causa a nivel mundial (4).

En Estados Unidos el sobrepeso y la obesidad son la segunda causa de muerte evitable, siendo el aumento del índice de mortalidad proporcional al aumento de la obesidad. Así un hombre con un IMC mayor de 45 kg/m² puede llegar a perder 13 años de vida, mientras que uno moderadamente obeso puede acortar su esperanza de vida de 2 a 5 años (33).

La obesidad es un importante factor de riesgo de patologías metabólicas como la diabetes mellitus de tipo 2 y el síndrome metabólico, así como de enfermedades respiratorias como asma o apnea del sueño, enfermedades cardiovasculares como el síndrome coronario agudo y la insuficiencia cardíaca, y numerosos tipos de cáncer (39).

A continuación, se realizará una breve descripción de las principales patologías asociadas a obesidad:

a) Diabetes mellitus tipo 2

La diabetes mellitus consiste en una hiperglucemia crónica, que puede deberse a una insuficiente producción de insulina (diabetes tipo 1), o a una deficiente respuesta a una correcta de producción de insulina por la que la glucosa no se internaliza en los tejidos (diabetes tipo 2) (40). Actualmente, se conoce que el 87,5% de individuos con diabetes mellitus de tipo 2 diagnosticada son obesos (40). El incremento de peso aumenta la hiperinsulinemia y la resistencia a la insulina. Se ha descrito que la grasa abdominal parece estar más ligada a la resistencia a la insulina con respecto a la grasa en otras localizaciones del cuerpo. La resistencia a la insulina producida por la obesidad no siempre determina la aparición de una diabetes clínica, de hecho, la mayoría de los obesos no son diabéticos, por lo que debe existir algún otro factor para que aparezca la

enfermedad. Lo que sí ocurre es que, por el contrario, el 80% de los diabéticos son obesos, por lo que constituye una comorbilidad de gran trascendencia a tener en cuenta. En cualquier caso, se sabe que la pérdida de peso mejora el control y el pronóstico de la enfermedad y aumenta la sensibilidad a la insulina (33).

b) Enfermedades Cardiovasculares

Hay muchas evidencias clínicas y epidemiológicas que relacionan la obesidad con diferentes tipos de enfermedades cardiovasculares, como la enfermedad coronaria, el infarto agudo de miocardio, la hipertensión, enfermedad cerebrovascular, fibrilación auricular, arritmias ventriculares y muerte súbita (41,42).

La obesidad puede aumentar la morbimortalidad de enfermedades cardiovasculares directamente, al adaptar la estructura y la función del sistema cardiovascular al exceso de peso corporal, mediante efectos adipocinéticos en la homeostasis vascular, que pueden participar en la génesis de la trombosis e indirectamente participan otros factores de riesgo cardiovascular concomitantes como son la resistencia a la insulina, la diabetes mellitus tipo 2, la adiposidad visceral, la hipertensión y la dislipidemia (41,42).

Este aumento de la morbimortalidad se debe entre otros aspectos a que cuando existe un balance energético positivo el exceso se acumula en el tejido adiposo que se expande. Cuando esta hiperplasia llega a su límite aumentan los ácidos grasos libres circulantes y comienzan a acumularse lípidos en zonas ectópicas como pueden ser: el tejido adiposo visceral, intrahepático, intramuscular, pericárdico, miocárdico, grasa perivascular, etc. (43).

Como resumen, en la tabla 3 se describen algunos efectos de la obesidad sobre distintas funciones cardíacas (42).

Tabla 3. Efectos de la obesidad a nivel cardiaco, basado en datos Ortega-Loubon et al (42).

NIVELES CARDIACOS	EFECTOS DE LA OBESIDAD
Hemodinámica	<ul style="list-style-type: none"> ➤ Aumento del volumen sanguíneo ➤ Aumento del gasto cardiaco ➤ Aumento de la presión arterial ➤ Aumento de la sobrecarga de la pared ventricular izquierda ➤ Hipertensión arterial pulmonar
Estructura cardiaca	<ul style="list-style-type: none"> ➤ Hipertrofia del ventrículo izquierdo ➤ Hipertrofia de la aurícula izquierda ➤ Hipertrofia del ventrículo derecho
Función cardiaca	<ul style="list-style-type: none"> ➤ Disfunción diastólica del ventrículo izquierdo ➤ Disfunción sistólica del ventrículo izquierdo ➤ Fallo del ventrículo derecho
Miocito	<ul style="list-style-type: none"> ➤ Esteatosis ➤ Hipertrofia ➤ Apoptosis ➤ Fibrosis

c) Cáncer

Existen múltiples evidencias de que el riesgo de sufrir cáncer aumenta en las personas que sufren sobrepeso u obesidad respecto a las personas que tienen un peso normal. Aproximadamente un 20% de todos los tipos de cáncer están asociados con el exceso de tejido adiposo, siendo así la obesidad uno de los principales riesgos de morbimortalidad evitables (44-46).

La Sociedad Americana de Cáncer junto con la Agencia para la Investigación del Cáncer Internacional y la Fundación de Investigación de Cáncer Mundial concluyeron que el cáncer de endometrio, adenocarcinoma de esófago, colorrectal, mama, próstata, y renal estaban estrechamente relacionados con la obesidad (44). Mientras que otros

cánceres como la leucemia, el linfoma no-Hodgkin, el mieloma múltiple, el melanoma maligno y los tumores de tiroides mantienen una relación menor con la obesidad (44). La mayor relación de causalidad que se ha encontrado con la obesidad fue con el cáncer de mama, estableciendo en un 80% de los casos (44). A nivel global, se estima que para las mujeres de Norte América, Europa y Oriente Medio un 9% de los cánceres diagnosticados están relacionados con la obesidad (39,47). Por ello, el control del peso de un individuo, evitando el sobrepeso y la obesidad, es uno de los principales factores de prevención para la reducción del riesgo de cáncer de mama, colon, útero, riñón y esófago (39).

No se sabe exactamente cuál es el mecanismo que relaciona el exceso de grasa con el desarrollo del cáncer, pero existen múltiples factores que pueden contribuir potencialmente a esta relación, por ejemplo, la obesidad se relaciona con defectos metabólicos, producción de adipocitoquinas inflamatorias, estrés oxidativo, dislipemia e hiperinsulinemia con resistencia periférica a la insulina (45).

Actualmente se reconoce el tejido adiposo como un órgano endocrino activo que secreta entre otras moléculas las adipocinas. Algunas de las funciones de la adiponectina son regular el metabolismo de la glucosa y los lípidos, la sensibilización a la insulina y efectos de regulación del sistema inmune, por lo que la hipoadiponectinemia está relacionada entre otros con varios tipos de cáncer (44,45).

d) Enfermedades respiratorias

La obesidad tiene efectos deletéreos en la función pulmonar, los mecanismos respiratorios, el intercambio de gases, el control de la respiración y la capacidad de realizar ejercicio físico. También se sabe que las apneas del sueño obstructivas, el asma, las enfermedades pulmonares obstructivas crónicas y las embolias pulmonares están relacionadas con la obesidad (48).

La apnea del sueño obstructiva consiste en episodios repetidos de obstrucciones totales o parciales de la vía aérea superior durante el sueño, lo que produce una reducción del flujo de aire. La obesidad se considera el principal factor de riesgo para su desarrollo, ya que la deposición ectópica de tejido adiposo en las vías respiratorias favorece el colapso durante el sueño (49). La apnea del sueño obstructiva también

umenta el riesgo de enfermedades renales y cardiovasculares debido a la reducción de actividad física por el cansancio, y debido a diferentes procesos metabólicos y hormonales (49).

Existen numerosos estudios que relacionan la obesidad con el asma, y aunque la mayoría de ellos utilizan el IMC, que no diferencia entre masa muscular y masa grasa, existe una relación entre el incremento de la grasa abdominal y el asma (50-52). Los datos apoyan la hipótesis de que la obesidad parece orientar a los pacientes hacia un asma con dominio de inflamación. Otros aspectos descritos como participantes en la relación entre el asma y la obesidad son el déficit de vitamina D, procesos de inflamación generalizada y el papel del microbioma intestinal (51).

1.4.- CONCEPTOS BÁSICOS DE GENÉTICA

a) La información genética (Cromatina, cromosomas y gen)

Los avances en tecnología han permitido avanzar mucho en el conocimiento de la genética desde que James Watson y Francis Crick descifrarán el código genético y la estructura del ADN en 1953 (53). En el núcleo de las células de los humanos se encuentra codificada la información genética. La forma de almacenamiento de esta información es la cromatina. La cromatina puede adquirir la forma de máximo plegamiento, generando diferentes unidades contenedoras de la información, denominados cromosomas. Los humanos tienen 22 pares de cromosomas autosómicos y 1 par de cromosomas sexuales, en total 23 pares de cromosomas, heredándose un cromosoma de cada par por progenitor (53).

El ADN (ácido desoxirribonucleico) es una molécula formada por una cadena de nucleótidos. Los nucleótidos contienen un azúcar (desoxirribosa), un grupo fosfato para unir los diferentes nucleótidos entre sí y una base nitrogenada que puede ser adenina (A), timina (T), citosina (C) o guanina (G). El ADN realmente se presenta en forma bicatenaria, es decir, formando una doble hélice con dos cadenas de nucleótidos que se unen entre sí por pares complementarios (A-T y C-G) mediante uniones no covalentes, puentes de hidrógeno (53). Por ello, se dice que ambas cadenas son complementarias y, por tanto, tienen la misma información, pero recogida de manera inversa y reversa.

La información genética viene contenida en la secuencia lineal de las cuatro bases A, T, G y C en cada cadena de ADN.

Los genes son las regiones de la molécula de ADN que contienen la información necesaria para que a través de los procesos de transcripción y traducción se puedan sintetizar las proteínas. De forma muy breve, la transcripción consiste en la generación de ARN a partir de una cadena de ADN. El ARN tiene una estructura monocatenaria y la secuencia de nucleótidos es idéntica a la cadena de ADN que ha utilizado como molde a excepción de que se sustituye la base nitrogenada timina del ADN por un uracilo (U) en el ARN. Una vez generado el ARN se produce la traducción en los ribosomas, que consiste en leer la información codificada en las secuencias de nucleótidos en ventanas de lectura de 3 nucleótidos o codones para generar un aminoácido, es decir, cada tres nucleótidos se generará un aminoácido. El código genético define los aminoácidos que corresponde a cada triplete de nucleótidos o codón. Es importante remarcar que es un código degenerado, es decir, un aminoácido puede ser generado a partir de uno o más codones. También existe un codón de inicio de traducción y hasta tres codones distintos de finalización de traducción (véase tabla (53)).

Tabla 4. Código genético

AMINOÁCIDOS	SÍMBOLOS		TRIPLETES CODIFICANTES/ CODONES					
Metionina	Met	M	AUG (<i>triplete de iniciación</i>)					
Alanina	Ala	A	GCA	GCC	GCG	GCU		
Glicina	Gly	G	GGA	GGC	GGG	GGU		
Valina	Val	V	GUA	GUC	GUG	GUU		
Leucina	Leu	L	UUA	UUG	CUA	CUC	CUG	CUU
Isoleucina	Ile	I	AUA	AUC	AUU			
Prolina	Pro	P	CCA	CCC	CCG	CCU		
Fenilalanina	Phe	F	UUC	UUU				
Triptófano	Trp	W	UGG					
Cisteína	Cys	C	UGC	UGU				
Ácido aspártico	Asp	D	GAC	GAU				
Ácido glutámico	Glu	E	GAA	GAG				
Arginina	Arg	R	AGA	AGG	CGA	CGC	CGG	CGU
Lisina	Lys	K	AAA	AAG				
Histidina	His	H	CAC	CAU				
Asparagina	Asn	N	AAC	AAU				
Glutamina	Gln	Q	CAA	CAG				
Serina	Ser	S	AGC	AGU	UCA	UCC	UCG	UCU
Treonina	Thr	T	ACA	ACC	ACG	ACU		
Tirosina	Tyr	Y	UAC	UAU				
CODONES STOP	STOP	X	UAA	UAG	UGA (<i>tripletes de stop</i>)			

Aminoácidos no polares.
 Aminoácidos polares cargados negativamente

Aminoácidos polares cargados positivamente.
 Aminoácidos polares no cargados

Tripletes STOP, no codificantes para aminoácidos.

En esta tabla se muestran los 61 tripletes o codones codificantes para los 20 aminoácidos esenciales y el triplete de inicio de traducción. También se muestran los tres tripletes de finalización de traducción o señal STOP.

b) Variabilidad genética (alteraciones genéticas)

Una alteración genética puede definirse como cualquier variación en el orden secuencial de los nucleótidos de la molécula de ADN. Sin embargo, no todas las

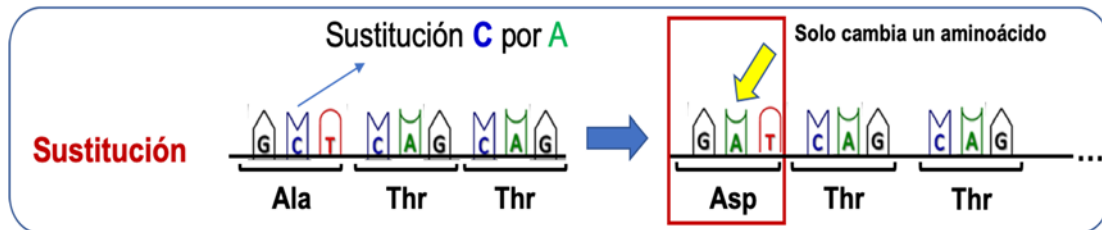
alteraciones genéticas tendrán la misma repercusión sobre la salud del portador. En función de la prevalencia en la población, se puede distinguir dos tipos de alteraciones genéticas:

- **Mutación** es una variación en la secuencia genética que conlleva cambios importantes en la funcionalidad y/o expresión de la proteína a la que codifica el gen. Es un tipo de alteración genética poco frecuente, afectando a un porcentaje de población inferior al 1%, y suele asociarse a efectos adversos sobre la salud del portador.
- **Polimorfismo**, también es una alteración genética como en el caso de la mutación, pero el polimorfismo tiene una prevalencia superior al 1%. Los polimorfismos suelen asociarse a cambios moderados en la funcionalidad y/o expresión de la proteína, teniendo efectos que pueden pasar desapercibidos (silentes) o bien generar algún tipo de ventaja o desventaja al portador. Cuando la alteración genética consiste en una sustitución de un único nucleótido se denomina SNP, del inglés Single Nucleotide Polimorphism. Si ocurre en la secuencia del exoma puede dar lugar a cambios en la funcionalidad de la proteína; mientras que si ocurre en regiones promotoras puede generar cambios en la expresión proteica (27,53). Se ha calculado que los SNPs aparecen de media cada 100-300 nucleótidos a lo largo del genoma. El genoma humano está compuesto por más de 3.000 millones de bases, por lo que un individuo podría tener millones de SNPs. La presencia de diferentes polimorfismos o variantes genéticas en los distintos genes es lo que va a determinar las características específicas de cada persona, incluyendo sus características para poder mantener mejor el peso o su capacidad para quemar grasas durante el ejercicio físico.

Existen distintos tipos de mutaciones o polimorfismos, sin embargo, la alteración más frecuente y la que se va a hacer constantemente referencia en este trabajo de tesis doctoral es la sustitución. La sustitución consiste en el intercambio o reemplazo de un nucleótido por otro, esta permutación de nucleótidos no conlleva cambios en las ventanas de lectura de traducción, por lo que en la mayoría de los casos únicamente se verá afectado un único aminoácido de la proteína (Figura 2). Es importante recalcar que

en el caso de que esa sustitución genere un codón de STOP de traducción se generará una proteína truncada con una estructura y funcionalidad altamente afectada (Figura 2).

Figura 2. Alteración genética. La sustitución de un nucleótido



En la sustitución se sustituye un nucleótido por otro, en la inserción se incluye un nucleótido. Como cada tres nucleótidos se formará específicamente uno de los aminoácidos de la proteína, la sustitución normalmente afecta a un único aminoácido, exceptuando los casos en los que se generen un codón o triplete de STOP, que generará una proteína incompleta o truncada

c) Nomenclatura de los polimorfismos de un único nucleótido (SNP)

Las alteraciones genéticas se deben nombrar siguiendo una nomenclatura específica, como en este trabajo únicamente se estudian SNPs nos vamos a centrar en la nomenclatura específica para las sustituciones.

Las variantes genéticas se pueden nombrar desde el cambio de nucleótidos en el gen o bien desde el cambio a nivel de aminoácidos en la proteína generada:

- **Desde el nucleótido en el gen:** el esquema generalmente utilizado para describir una variante genética es el nombre del nucleótido original seguido de la posición donde ocurre el cambio y después el nombre del nucleótido <<menor>>. Por ejemplo, en el gen AGRP, G105T significa que el nucleótido guanina (G) se ha sustituido por una timina (T) en la posición 105 del gen AGRP.
- **Desde el aminoácido en la proteína:** el esquema utilizado para describir una variante a nivel proteico es el "aminoácido habitual o control + localización del aminoácido en la proteína + aminoácido nuevo". Por ejemplo, la alteración en la proteína leptina R308D, significa que el aminoácido arginina (R) en posición 308 se sustituye por el aminoácido aspártico (D). La R y la D son las abreviaturas internacionales que corresponden a la arginina (R) y el

aspártico (D). También se podría utilizar la fórmula Arg308Asp para nombrar este polimorfismo.

1.5.- IMPACTO DEL FACTOR GENÉTICO EN EL SOBREPESO Y OBESIDAD

Diferentes evidencias experimentales han puesto de manifiesto que los factores genéticos están directamente relacionados con el 30-40% de los casos de obesidad, no sólo en las formas monogénicas, sino también en la obesidad común (54). Lamentablemente, aunque en los últimos años se están realizando importantes progresos sobre las características genéticas de diferentes patologías, el control genético de las formas comunes de obesidad en el humano no se comprende bien (54).

En la actualidad, la contribución de los factores genéticos a la obesidad se puede clasificar en tres grandes grupos según su etiología:

- **Obesidad monogénica no sindrómica:** es la causa de aproximadamente el 5% de los casos de obesidad. Esta forma de obesidad se debe a mutaciones localizadas en un único gen, que se hereda en forma mendeliana. Este tipo de obesidad se asocian con fenotipos muy severos a edades muy tempranas.

En la actualidad se conocen más de 200 alteraciones genéticas identificadas en unas decenas de genes que se han relacionado con la obesidad monogénica. Entre los genes causantes de obesidad monogénica, destacan los siguientes:

- El **gen FTO** es el que parece tener una mayor relación con la obesidad, aunque de forma directa únicamente influye en el 1% del peso corporal. El gen FTO codifica una proteína nuclear con actividad de dimetilasa de ácidos nucleicos. Entre las acciones que desempeña destacan la reparación del ADN, la homeostasis de la temperatura y la regulación del almacenamiento de lípidos en el tejido adiposo. Los niveles de esta proteína se regulan en el hipotálamo con la ingesta y el ayuno. Se han encontrado al menos 20 polimorfismos en regiones intrónicas que influyen de algún modo sobre la obesidad, y también sobre la diabetes mellitus tipo 2 y el síndrome metabólico (27).

- El **gen LEP** (leptina) codifica la hormona leptina. La leptina se sintetiza en los adipocitos y su síntesis depende de varios factores como la liberación de insulina durante el consumo de alimentos, pero también de factores como el estrés y la práctica de ejercicio físico. La leptina está estrechamente ligada al riesgo de obesidad, ya que a través de su unión al receptor de leptina manda la señal de saciedad al cerebro reduciendo la ingesta de alimentos. Además, la leptina también actúa sobre el músculo esquelético, promoviendo el gasto energético a través de la degradación de ácidos grasos y, por tanto, reduciendo el riesgo de infiltración de grasa muscular (27). Un ejemplo de alteración genética en leptina asociada a obesidad monogénica, es la sustitución C>T en la posición 105, que convierte un codón de arginina en un codón de terminación prematura y que no permite la unión a su receptor, lo que conlleva una continua necesidad de ingerir alimentos¹⁹. Posteriormente, se demostró que la administración exógena de leptina generaba una disminución de ingesta y una reducción de obesidad (55-57).

- El **gen MC4R** codifica el receptor de melanocortina 4 y también, como en el caso de la leptina, controla el apetito y la saciedad desde el hipotálamo regulando así la homeostasis energética. Desencadena efectos anorexigénicos, mediante su unión a las melanocortinas. Esta molécula se sintetiza en un tipo específico de neuronas del hipotálamo, bajo el estímulo de la insulina y la leptina. Sus mutaciones son la causa genética más frecuente de obesidad en humanos, habiéndose descrito más de 150 mutaciones relacionadas con obesidad monogénica. Estas mutaciones representan casi el 10% de la obesidad heredable grave (27).

- El **gen AdipoQ** (adiponectina) codifica a la hormona adiponectina que regula el metabolismo energético del organismo y se produce en los adipocitos. Tiene el papel de estimular la oxidación de ácidos grasos, disminuir la concentración de los triglicéridos plasmáticos y facilitar el metabolismo de la glucosa al aumentar la sensibilidad a la insulina. Una reducción de la concentración de adiponectina plasmática se ha asociado con obesidad,

diabetes tipo 2 y una mayor frecuencia de complicaciones cardiovasculares (27). Un claro ejemplo es la mutación en la posición Cys39 generada en ratones, que genera una muy reducida expresión de adiponectina y produce unos ratones obesos que tiene un peso cuatro veces mayor que los wild-type (58).

- **Obesidad sindrómica:** se refiere a aquellos casos en los que la obesidad moderada o extrema es una característica más del conjunto de las características clínicas que conforma el síndrome. La causa es siempre genética y está asociada principalmente a alteraciones cromosómicas u otras formas más complejas de herencia, como dato adicional, suelen cursar con retardo mental. En la actualidad, se conocen más de 25 causas sindrómicas obesogénicas (59)

-**Obesidad poligénica:** aquella causada por la interacción de diferentes variantes genéticas que interactúan entre sí y con los hábitos conductuales y características ambientales a las que está expuesta cada individuo. Es decir, cada alteración aportaría un efecto sobre el peso corporal y su contribución sería modulada por los diferentes factores modificables y no modificables expuestos anteriormente (*véase sección 1.- Etiología clásica del sobrepeso/obesidad. Factores asociados*). Este tipo de obesidad es el más frecuente en la población, sin embargo, es el menos conocido y es uno de los principales focos de atención de la sociedad médico-científica.

En el mapa genético de la obesidad humana (versión 2005) se determinan al menos 127 genes candidatos como implicados en la obesidad poligénica, donde se han reconocido más de 600 loci (35). Sin embargo, las evidencias experimentales parecen señalar unas decenas de genes donde la presencia de alteraciones genéticas interacciona entre sí y con aspectos ambientales para modular el riesgo de obesidad de los portadores. En este sentido, se ha señalado la existencia de polimorfismos SNPs en 8 genes como de particular interés en la adipogénesis y lipogénesis humana: PPARG, INSIG2, ADRB2, ADRB3, UCP2, AGRP, LEP y el LEPR. Como ejemplo, en el estudio de Garaulet M. et al. se describió que los portadores del alelo Ala12 en el gen PPARG2 mantenían mejor el peso, siendo más resistentes tanto a la pérdida de peso como a la

ganancia, en comparación con los portadores del alelo Pro12 (60). También cabe citar el estudio de Kaulfers A. M. et al., en el que se encontró que existía relación entre el polimorfismo rs7566605 del gen INSIG2 y el metabolismo de los lípidos, asociándose con sobrepeso y obesidad (61). Otro ejemplo podría ser el estudio de Nicoletti C. F. et al. en el que se estudiaron los polimorfismos Ala55Val y -866G>A del gen UCP2, encontrándose una reducción de la formación de ATP a partir de ADP, que incluso afectó a la liberación de insulina. Este estudio encontró que los individuos portadores del alelo T para el polimorfismo Ala55Val perdían peso más fácilmente (62).

En la última década, gracias al imparable avance en el análisis de genotipado simultáneo, se ha incrementado el número de estudios basados en la técnica de asociación del genoma completo (GWAS). Estos estudios han identificado cientos de SNP relacionados con variaciones en el IMC (32,53,64). Sin embargo, la capacidad de predicción de los GWAS se ha puesto en duda por diferentes investigadores, ya que la aproximación básica de GWAS es evaluar la asociación entre cada marcador de genotipo y un fenotipo de interés entre un gran número de individuos. Es decir, cada individuo, o mejor dicho, su genoma es relacionado con su fenotipo. Esto hace que, la técnica de GWAS no tenga capacidad de predicción de un SNP para un fenotipo. No obstante, la estrategia GWAS, es capaz de explicar menos del 2% de las variaciones de IMC entre distintos individuos, y no llega a explicar ni un 1% de la relación en el diámetro cintura-cadera con sobrepeso y obesidad entre individuos diferentes (65,66). Es interesante, que la epistasis genética, que se define como una relación de enmascaramiento entre dos locus, es decir, la interacción entre ellos ocurre cuando el efecto de uno de los locus es dependiente de los genotipos del otro de los locus, podría tener un papel relevante a la hora de poder predecir la probabilidad genética que presenta un individuo para desarrollar sobrepeso u obesidad (67).

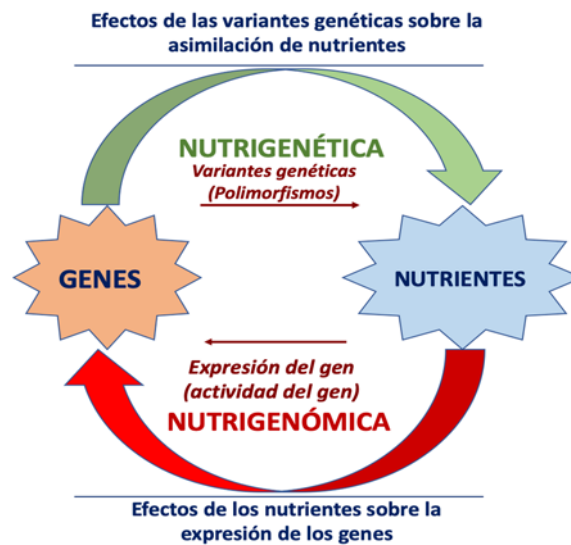
Sin embargo, si se analiza la literatura científica existente sobre la genética de la obesidad se observan un gran número de asociaciones incongruentes. Este hecho puede deberse a la heterogeneidad de las personas estudiadas, con diferentes grupos étnicos y/o grados de obesidad (leve en algunos estudios y grave en otros). Todas estas características sugieren la necesidad de un abordaje metodológico más preciso en el que se puedan analizar de forma conjunta las variantes genéticas, los factores ambientales

e incluso hábitos conductuales de los individuos para ver el efecto conjunto sobre el riesgo de obesidad. Como alternativa, los sistemas de inteligencia artificial en el que se establecen algoritmos predictivos de aprendizaje en base a las características de los individuos, parecen aportar una novedosa y esperanzadora línea de investigación.

1.6.- DEFINICIÓN DE NUTRIGENÉTICA. GENES RELACIONADOS CON SOBREPESO/OBESIDAD.

La nutrigenética es la ciencia que se encarga de estudiar e identificar aquellas variantes genéticas (polimorfismos o mutaciones) que podrían aportar a la persona ventajas o desventajas en la asimilación de los diferentes componentes de la dieta y, por tanto, predecir el riesgo de obesidad individualizado, en función del perfil genético heredado (68,69). A través de la nutrigenética se pueden identificar los nutrientes de la dieta que se metabolizan mejor o peor y utilizar este conocimiento para el control del peso. Por otro lado, los factores ambientales también tendrían una gran importancia en el nivel de activación o inhibición de cualquier gen, y que esto tenga por lo tanto influencia a su capacidad funcional. Es decir, no sólo será importante la variabilidad genética de una persona para conocer cómo puede metabolizar los diferentes componentes de la dieta, sino también sus hábitos de vida.

La nutrigenética cuenta con un amplio abanico de posibilidades en sus aplicaciones. La nutrigenética no sólo puede aplicarse con fines meramente estéticos, sino, que además la identificación de determinadas variantes en los genes relacionados con la nutrición también permite predecir el riesgo de sufrir futuros problemas de salud relacionados con la obesidad.

Figura 3. Diferencias entre nutrigenética y nutrigenómica.

El control del ciclo hambre-saciedad, el metabolismo de carbohidratos y grasas, así como el efecto del ejercicio físico y la dieta sobre la pérdida de peso son aspectos cruciales que determinarán en gran medida el riesgo de sobrepeso/obesidad. En la actualidad, se conocen numerosas variantes genéticas en genes que modulan los procesos anteriormente descritos. A continuación, se describen los principales polimorfismos en cuanto a evidencia científica relacionada con cada uno de los aspectos estrechamente ligados a sobrepeso/obesidad.

1.6.1.- Ciclo saciedad-hambre

El ciclo hambre-saciedad tiene un papel fundamental en el control del peso corporal ya que regula la relación entre energía ingerida y energía consumida. Este ciclo lo podemos dividir en tres etapas principales:

- *Etapa del apetito*, relacionada con las señales que inician la ingesta de alimentos.
- *Etapa de la satisfacción*, relacionada con las señales de terminación de la ingesta.
- *Etapa de la saciedad*, modulada por señales que inhiben la ingesta entre comidas.

Hay dos indicadores que regulan el ciclo hambre-saciedad: a) la respuesta hormonal y b) la respuesta mecánica, debida a la distensibilidad del estómago y del duodeno.

Figura 4. Fases ciclo saciedad-hambre.

a) Control hormonal del ciclo saciedad-hambre

La necesidad de comer ocurre fundamentalmente como una consecuencia de la depleción energética y viene regulada por múltiples mediadores neurohormonales. Dentro del hipotálamo existen dos tipos de neuronas que producen señales orexigénicas (inductoras del apetito) y las anorexigénicas (inhibidoras del apetito o señales de saciedad). Estas dos vías se han propuesto como las de mayor importancia en la patogénesis de la obesidad.

Entre los genes relacionados a las señales orexigénicas se encuentran el gen del neuropéptido Y (NPY), la proteína relacionada a Agouti (AGRP), ghrelina y el receptor de ghrelina (GHSRs). Los genes relacionados a señales anorexigénicas son también diversos, entre los que se encuentran la leptina y su receptor (LEPR), la pro-opiomelanocortina (POMC), el péptido YY (PYY), la insulina y los receptores de las melanocortinas (MC3R y MC4R).

Los dos conjuntos de neuronas orexigénicas y anorexigénicas reciben información de varias hormonas endocrinas de la siguiente manera. La leptina se secreta a partir del tejido adiposo, circula a niveles proporcionales a las reservas de grasa corporal y ejerce sus efectos a través del receptor de leptina (LEPR), inhibiendo las neuronas NPY / AGRP y estimulando las neuronas POMC / CART. La glucosa también puede ser un regulador

de leptina, ya que existe una relación entre la reducción del metabolismo de la glucosa en los momentos de ayuno y la reducción de la liberación de leptina. Sin embargo, al aumentar el metabolismo de glucosa en el periodo inmediato postprandial aumentan los niveles de leptina (70). El páncreas segrega insulina, que tiene una influencia anorexigénica sobre el núcleo arqueado. La grelina es producida por el estómago y el duodeno y estimula las neuronas NPY / AGRP a través de sus receptores de la hormona del crecimiento (GHSR). El péptido YY3-36 (PYY3-36) se secreta desde el tracto gastrointestinal distal y envía señales a través de los receptores Y2 (Y2R) para producir un efecto inhibitor sobre las neuronas NPY / AGRP. Las neuronas NPY / AGRP también tienen un efecto inhibitor sobre las neuronas POMC / CART a través de la liberación de ácido γ -aminobutírico (GABA), que podría ser estimulado por la unión de grelina a GHSR. Las señales orexigénicas y anorexigénicas que son producidas por las neuronas NPY / AGRP y POMC / CART son dirigidas a las neuronas efectoras de segundo orden, que también reciben entradas modificadoras de las señales de dopamina y serotonina. Estas neuronas efectoras expresan receptores que incluyen el receptor Y1 (Y1R) y el receptor de melanocortina 4 (MC4R). Estos procesos se unen para proporcionar el equilibrio general entre la ingesta de alimentos y el gasto energético.

Otro grupo de genes relacionados a señales periféricas y no centrales, favorecen la adipogénesis como el gen de la adiponectina (ADIPOQ) y el receptor gama 2 del proliferador activado de los peroxisomas (PPARG2), o, por el contrario, producen una mayor destrucción de adipocitos como los receptores adrenérgicos beta 2 y 3 (ADRB2 y ADRB3) y las proteínas de desacoplamiento mitocondrial (UCP2 y UCP3) entre muchos otros.

b) Control mecánico del ciclo saciedad-hambre

La composición de los alimentos y nutrientes van a influir directamente sobre el vaciamiento gástrico y, por tanto, en los niveles de saciedad y en la ingesta de alimentos. Menor vaciamiento y un mayor volumen genera distensión en el intestino que provoca la liberación de la colecistoquinina (hormona asociada a la sensación de saciedad). Por ejemplo, la fibra soluble proporciona sensación de saciedad. Por ello, el consumo de fibra soluble conlleva una menor ingesta energética favoreciendo el mantenimiento del

peso o su pérdida. Por otra parte, las proteínas y grasas también retrasan el vaciamiento gástrico respecto los hidratos de carbono La fibra insoluble, que aporta volumen a las heces y parece que ayuda a que los alimentos pasen más rápidamente a través del estómago e intestino siendo esto beneficioso para la prevención del cáncer de colon al facilitar la eliminación rápida de toxinas y sustancias que pueden producir cáncer.

c) **Genética del ciclo saciedad-hambre**

Por ello, aquellas alteraciones genéticas que afecten a la funcionalidad o la regulación de alguno de los genes anteriormente descritos que juegan un papel crucial en el control del ciclo saciedad están estrechamente asociadas con alto riesgo de sobrepeso/obesidad:

En este sentido, una sobreexpresión del **gen AgRP**, que actúa como mediador en la regulación del balance energético con efecto orexígeno, conduce a un aumento de ingesta de alimentos y, por tanto, favorece la obesidad. De hecho, se encuentran mayores niveles de AgRP plasmático en personas obesas que en personas sin sobrepeso (71). En el gen AgRP se ha identificado también un polimorfismo Thr67Ala (rs5030980) que parece estar asociado con la anorexia nerviosa, aunque según diferentes estudios parece que depende de la raza, ya que en algunos de ellos no se ha encontrado relación con la obesidad, aunque también se podría deber a diferencias en la estadística o a la muestra de población elegida (71-73).

El gen **MC4R**, es un punto clave en la regulación del ciclo saciedad-hambre, como se ha comentado anteriormente. En el gen MC4R destacan dos polimorfismos por el conocimiento sobre sus efectos en el riesgo de obesidad, el rs17782313 y el rs571312. La variante genética del rs17782313 cambia el alelo C por un alelo T. El alelo C de este polimorfismo se ha relacionado con una menor actividad del MC4R y, por lo tanto, una mayor ingesta y un mayor riesgo de obesidad (74). En estudios recientes, se ha demostrado que los individuos que presentaban el alelo C en el polimorfismo rs17782313 tenían unos niveles mayores de grelina en situación postprandial con una mayor prevalencia de comer en exceso, y esta asociación era mayor en mujeres que en hombres (75). El polimorfismo rs571312, consistente en un cambio del alelo C por el alelo A se ha asociado a un mayor IMC tanto en adultos como en adolescentes y niños (76).

En el **gen LEP**, se han descrito diferentes polimorfismos asociados con variación en los niveles de leptina en suero (77,78). El polimorfismo más estudiado en este gen es el rs7799039 basado en un cambio del nucleótido G por el A en la región 5' no traducida del gen (-2548G>A) que se ha relacionado con la obesidad. Los primeros estudios realizados asociaron este SNP con niveles bajos de leptina (79,80), obesidad (80,81) y aumento del IMC (82).

En el **gen LEPR**, también se han descrito la presencia de variantes polimórficas que predisponen a una disminución en la funcionalidad del receptor de la leptina y, por tanto, un papel crucial en el desarrollo de la obesidad humana. En este sentido, en un estudio reciente, Sun y colaboradores realizaron un estudio GWAS en 1504 mujeres de origen europeo (83). De entre todos los SNPs localizados, destacaron dos: rs1137100 (326A>G) y rs1137101 que se asociaron con obesidad independientemente de los niveles de leptina sérica (83).

En el **gen de la adiponectina** se han identificado diferentes polimorfismos asociados con obesidad (84). Sin embargo, la variante más estudiada es la rs17300539 que consiste en el cambio del nucleótido G por en el promotor del gen (-11391 G>A) se halla en el promotor del gen de la adiponectina. Un estudio realizado en población española concluyó que la presencia del alelo A se asoció a una protección frente al incremento de peso (85).

1.6.2.- Metabolismo de los carbohidratos

Los carbohidratos constituyen la principal fuente de energía para el ser humano, es por ello por lo que entre el 50-60 % de la energía proporcionada por la dieta debería ser aportada por este nutriente. Nutricionalmente podemos diferenciar dos categorías:

- 1) los *carbohidratos glicémicos*, aquellos que son absorbidos y digeridos en el intestino delgado.
- 2) la *fibra dietética*, carbohidratos no digeridos que pasan al intestino grueso.

Sin embargo, la clasificación más común de los hidratos de carbono o carbohidratos se basa en su tamaño y complejidad, diferenciándose entre carbohidratos sencillos (monosacáridos y disacáridos) y carbohidratos complejos (almidón, glucógeno, fibra).

En la genética asociada al metabolismo de los hidratos de carbono, se han identificado también genes y polimorfismos genéticos que podrían tener un papel relevante:

- El **UCP2** es un gen localizado en el cromosoma 11 (región 11q13) que ha sido asociado con la obesidad y cambios de peso. Este gen codifica las proteínas de desacoplamiento, que se encuentran en la membrana interna de las mitocondrias permitiendo el transporte de protones y aniones, y así reduciendo la formación de ATP. Esto produce efectos de desacoplamiento en tejidos específicos como el cerebro, el riñón, el hígado y la piel entre otros, interviniendo en la termogénesis y en el gasto energético, reduciendo la formación de compuestos de oxígeno reactivo, regulando el metabolismo de ácidos grasos, de moléculas de ATP dependientes de las vías metabólicas, la temperatura corporal, la composición del cuerpo y el metabolismo de la glucosa. El gen UCP2 se expresa principalmente en el tejido adiposo y en los islotes pancreáticos, y está ampliamente asociado con la obesidad y cambios de peso, especialmente con el efecto de la obesidad en la sensibilidad de la insulina y la predisposición a la diabetes. Una sobreexpresión de este gen se traduce en un descenso de la producción de ATP y una defectuosa secreción de insulina (62,86,87).

El polimorfismo Ala55Val (alelo T) (rs660339) parece favorecer la pérdida de peso, las personas con genotipo CT y TT pierden peso de forma más efectiva que los de genotipo CC, y la mayor pérdida de peso la presentan aquellos con genotipo TT. Existen varios estudios en los que parece ser que el genotipo TT presenta un menor grado de desacoplamiento, un uso de la energía más eficiente, un aumento de la producción de compuestos de oxígeno reactivo y menor oxidación de grasas comparados con los que tienen genotipo CC. Parece ser que el gen UCP2 interviene en el ajuste del aumento del uso de lípidos cuando disminuye la glucosa disponible. Sin embargo, el efecto de este polimorfismo presenta diferencias según el género y raza (62,86).

- El **gen peroxisome proliferator-activated receptor γ (PPARG)** se localiza en el cromosoma 3 (3p25). Este gen interviene en la diferenciación de adipocitos, la oxidación de lípidos, la homeostasis de la glucosa, el metabolismo de hidratos de

carbono y los ácidos grasos, procesos inflamatorios, adipogénesis, resistencia a la insulina, hipertensión arterial, dislipidemias Y minimiza la grasa ectópica e intra-abdominal (27,88,89). El alelo Ala del polimorfismo Pro12Ala del gen PPAR γ (rs1801282) se asocia con un elevado IMC, especialmente entre individuos obesos (89). Esto se debe a que con este disminuye la actividad del receptor provocando una mayor sensibilidad de la insulina y aumenta la producción de lípidos (60). Existe una asociación entre el genotipo PPAR γ Pro12Ala y la concentración de insulina en sangre, las personas con genotipo Ala12 tienen menores niveles de insulina rápida que aquellos con el genotipo Pro12Pro, y también se observa que los portadores del alelo Ala12 con poca actividad física atenuaban la obesidad, aunque solo es significativo cuando la ingesta de ácidos grasos monoinsaturados es mayor del 56%. Según las observaciones los portadores del alelo Ala12 mantienen su peso más estable que los portadores del alelo Pro12, tanto frente a la ganancia como a la pérdida de peso. El alelo Ala12 muestra un rol protector frente a la resistencia a la insulina y parece tener una interacción con la ingesta de ácidos grasos monoinsaturados y el IMC (60,90). Se ha encontrado una asociación significativa entre este polimorfismo y los cambios en la composición corporal en pacientes obesos tras 12 semanas de intervención con aceite de oliva virgen extra. Los portadores del alelo Ala presentan un mayor descenso en el porcentaje de grasa corporal que los homocigotos Pro12Pro (89). Otro estudio sobre este polimorfismo mostró que tras un programa de entrenamiento aeróbico los individuos que presentaban la variante Pro12Pro habían conseguido una mayor pérdida de grasa corporal comparado con los portadores del alelo 12Ala (91).

1.6.3.- Metabolismo de las grasas

Las grasas constituyen uno de los nutrientes principales de la dieta, particularmente se obtienen de los alimentos de origen animal y vegetal. En la dieta diaria, si seguimos la dieta Mediterránea, el aporte de grasa supone alrededor del 30% del total de nutrientes que diariamente consumimos.

En la genética asociada al metabolismo de las grasas también se han identificado genes y polimorfismos genéticos. En este sentido, la mayoría de los autores coinciden en que genes codificantes para apolipoproteínas relacionadas con el transporte de grasas como la apolipoproteína A-II, (APOA2) y el gen de la apolipoproteína A-V (APOA5), así como en genes implicados en la formación de colesterol y metabolismo lipídico como el gen de la masa grasa y de la obesidad (FTO), y el gen inducido por insulina 2 (INSIG2) son los más relevantes en el riesgo de obesidad por deficiente metabolización lipídica.

- Uno de los primeros estudios realizados en el **gen de la apolipoproteína A-II (APOA2)**, indicó que los individuos homocigotos para la variante “-265T>C” localizada en el promotor del gen APOA2 (polimorfismo rs5082) mostraban aumentadas significativamente los porcentajes de masa grasa reflejándose en un aumento de peso corporal, el IMC y la circunferencia de la cintura. Además, en estos individuos se observó un mayor consumo de alimentos compuestos por grasa y proteína, en comparación con los individuos homocigotos normales o heterocigotos para esta variante del gen (92).
- **La APOA5** es clave en la modulación del metabolismo de los triacilgliceroles (TG) y también puede modificar los niveles de otros componentes del metabolismo altamente relacionados con los triacilgliceroles, influyendo el proceso de síntesis y secreción de partículas ricas en TG. Este gen se localiza en el cromosoma 11q23, donde se han descrito distintos polimorfismos, muchos de ellos formando haplotipos. Los dos polimorfismos más estudiados son el polimorfismo -1131T>C, en el promotor del gen, y el polimorfismo 56C>G. Múltiples estudios han relacionado estos polimorfismos con mayores concentraciones plasmáticas de triglicéridos (93) y su participación en el almacenamiento y movilización de los lípidos intracelulares (94). Recientemente, en participantes en el estudio de Framingham se ha descrito una interacción entre el polimorfismo -1131T>C en el promotor del gen de la APOA5 y el riesgo de obesidad en función de la grasa consumida.
- El **gen INSIG2 (gen inducido por insulina 2)** que participa en dos vías implicadas en la regulación de la síntesis de colesterol, ácidos grasos y triacilgliceroles. El INSIG2 es un gen relacionado con el metabolismo lipídico,

principalmente debido a su papel inhibitorio de la síntesis endógena de colesterol y ácidos grasos. Por ello, todos aquellos polimorfismos que modifique su capacidad funcional sería esperable que indujeran un aumento de los niveles lipídicos plasmáticos y, por tanto, el riesgo de acumulación de grasa. En 2006, en 1.071 participantes del Estudio Framingham, se encontró que la variante CC del polimorfismo rs7566605 del gen INSIG2 tenía una asociación significativa con la obesidad (95,96) En este sentido, el Instituto Europeo de Biomedicina reportó que la presencia de la variante CC de este polimorfismo se asociaba con un 40% mayor de probabilidad de desarrollar obesidad respecto a los no portadores de CC.

- En casi todos los trabajos realizados hasta la fecha, la asociación más fuerte entre obesidad y gen, aparece insistentemente el **gen llamado FTO (fat mass and obesity associated, gen asociado a la masa de grasa y la obesidad)**. Las bases de datos de la literatura científica recogen más de 2.000 artículos sobre este gen publicados en los últimos años. Este gen fue identificado por primera vez como factor de riesgo para la obesidad por Frayling y colaboradores en el año 2007 (97) El gen FTO se localiza en el cromosoma 16 (16q12.2) y se ha determinado que codifica la producción de una proteína nuclear con actividad de demetilasa de ácidos nucleicos. Su acción se ha relacionado con diversos procesos bioquímicos y fisiológicos, entre los que destacan la regulación del almacenamiento de lípidos y del tejido adiposo.

1.6.4.- Ejercicio físico-dieta. Genes que intervienen en la eficacia del ejercicio físico para la mejora de pérdida de peso.

Como se ha comentado anteriormente, el ejercicio físico es una saludable actividad muy positiva para prevenir la obesidad y favorecer la pérdida de peso. Sin embargo, no todos los individuos se benefician de la misma manera en la pérdida de peso de ese ejercicio físico. Un aspecto muy importante es el papel de las variantes genéticas en la respuesta al ejercicio físico y la degradación de tejido adiposo. Por ello, conocer las variantes genéticas que modulan los rasgos relacionados con la obesidad y ejercicio

físico tiene un gran valor ya que permitiría individualizar recomendaciones y programas de ejercicios para hacerlos más eficientes en la pérdida de peso.

Recientemente se describió que la presencia de polimorfismos en los genes que producen proteínas que actúan como receptores de hormonas adrenérgicas (ADR) se han identificado como modificadores genéticos del porcentaje de grasa corporal en respuesta al ejercicio físico. Existen dos tipos de receptores adrenérgicos o ADR (alfa y beta). Los receptores adrenérgicos alfa y beta reciben una serie de estímulos durante el ejercicio físico, que van a inducir la degradación de la grasa (91,98). Por ello, los polimorfismos localizados en la familia de ADR pueden condicionar la degradación de grasa asociada al desarrollo de ejercicio físico en distintas personas.

- El **gen ADBR3 (beta3-adrenergic receptor)** se expresa en el tejido adiposo y está relacionado con la regulación del metabolismo energético (21). Una baja actividad de este receptor puede provocar obesidad al disminuir su función en el tejido adiposo. El polimorfismo Trp64Arg (rs4994) se asocia con una tendencia al exceso de masa corporal, resistencia a la insulina y diabetes tipo 2. Un estudio de Maha Daghestani et al., obtiene unos resultados en los que encuentran una relación significativa entre la obesidad y este gen en población saudita, ya que los individuos con sobrepeso y obesidad presentaban una mayor frecuencia significativa del alelo Arg64 frente a los individuos con normopeso (99). Hay muchos estudios que muestran mayores IMC en portadores del alelo Arg64 entre participantes sedentarios, pero no encontraron diferencias entre participantes que realizaban ejercicio físico. Otros estudios muestran que mujeres con el alelo Arg64 combinando ejercicio y dieta perdían menos peso que las que no tenían ese alelo. En otros estudios se encontró que los portadores de Arg64 perdían más grasa tras un entrenamiento aeróbico durante 24 semanas (91). Hay un metaanálisis en que se ve que esta relación varía con la raza, así los individuos del este de Asia presentan una mayor frecuencia del alelo Arg64 frente a los caucásicos, y se obtiene una asociación significativa entre el aumento del IMC y los portadores del alelo Arg64 únicamente en los asiáticos, siendo esta relación no significativa en los caucásicos (98).

- El **gen ADRB2 (beta2-adrenergic receptor)** se localiza en 5q31-q32 y es un gen relacionado con la regulación del balance energético mediante la estimulación de la termogénesis y la movilización de lípidos en el tejido adiposo y en la lipólisis inducida por catecolaminas en el tejido muscular (22,100).

Algunos polimorfismos pueden reducir la lipólisis favoreciendo así la obesidad, en concreto, el polimorfismo Gln27Glu, (rs1042714), en el que los portadores del alelo Glu27 limitan la infrarregulación del gen afectando a la masa corporal. Hay un estudio en el que las mujeres portadoras del alelo Glu27 y que se mantenían activas en su tiempo libre tenían un peso mayor que las no portadoras, lo que sugiere que pueden ser más resistentes a la pérdida de peso (91). Otro estudio encontró entre un grupo de mujeres obesas, que los homocigotos Glu/Glu presentaban un mayor porcentaje de grasa y un menor porcentaje de masa magra (101). Aunque existen estudios en los que no se encuentra relación entre este polimorfismo y el sobrepeso y la obesidad (102).

1.7.- NUEVAS TECNOLOGÍAS DE MANEJO DE LA INFORMACIÓN (INTELIGENCIA ARTIFICIAL)

Gracias a la revolución tecnológica para generar datos, al desarrollo de herramientas de análisis de datos y a una nueva mentalidad de los científicos para compartir resultados se ha hecho posible el uso del Big-Data en el área biomédica y su análisis mediante inteligencia artificial (103).

1.7.1.- Big-Data

La definición de Big-Data está basada en las cuatro uves: volumen, variedad, velocidad y veracidad. De esta forma se puede definir el Big-Data como un gran volumen de datos muy variados a los que se puede acceder rápidamente (velocidad) y que son veraces (103).

El Big-Data tiene aplicación en numerosas áreas, pero en concreto, la información recopilada en el área biomédica se caracteriza por tener un gran volumen, ser compleja y estar bastante desorganizada. Dicha información se puede encontrar en forma de anotaciones, informes, imágenes, resultados de pruebas, etc., siendo así muy diversa.

Por lo que cumple con los requisitos de la definición de Big-Data. Para poder trabajar con ella, es importante que esté recopilada correctamente para poder formar parte del Big-Data y de este modo que pueda ser analizada y procesada para su uso (104).

Para que esto sea posible, el concepto de “FAIR (Findable, Accesible, Interoperable y Reusable) data” se publicó en 2016 y es un conjunto de características que deben cumplir los datos recopilados. De este modo, la información almacenada debe ser fácil de encontrar, estando bien referenciada y aportando toda la información necesaria para ello; también debe ser accesible, esto significa que se debe poder recuperar la información, siempre y cuando se tenga permiso para ello; para poder usarlos deben estar recopilados en un lenguaje y formato que sean compatibles, o ir acompañados de las aplicaciones necesarias para poder trabajar con ellos; para poder usarlos, es importante que esté especificada su procedencia así como los derechos de uso que se les quieran aplicar (103, 105).

Tabla 5. Características de los principios FAIR. (Tabla traducida del artículo de Mark D. Wilkinson et al.)(105).

Principios FAIR	
Findable (Encontrables)	<p>Asignar un identificador global único y persistente a los datos y metadatos.</p> <p>Describir los datos con gran cantidad de metadatos.</p> <p>Los metadatos deben incluir clara y explícitamente el identificador de los datos que describen.</p> <p>Los metadatos y datos se deben registrar o indexar en un recurso de búsqueda.</p>
Accessible (Accesibles)	<p>Los datos y metadatos se pueden recuperar por sus identificadores usando protocolos de comunicación estandarizados.</p> <p>Dichos protocolos tienen que ser abiertos, gratuitos e implementables de forma universal.</p> <p>El protocolo debe permitir procedimientos de autenticación y autorización cuando sea necesario</p> <p>Los metadatos deben permanecer accesibles, incluso cuando los datos ya no estén disponibles.</p>
Interoperable (Interoperable)	<p>Los datos y metadatos deben usar un lenguaje formal, accesible, que se pueda compartir y ampliamente aplicable para representar el conocimiento.</p>

	<p>Los datos y metadatos deben usar un vocabulario que siga los principios FAIR.</p> <p>Los datos y metadatos deben incluir referencias cualificadas a otros datos o metadatos.</p>
<p>Reusable (Reutilizable)</p>	<p>Los datos y metadatos se deben describir con una multitud de atributos precisos y relevantes.</p> <p>Los datos y metadatos se deben publicar con una licencia de uso de datos clara y accesible.</p> <p>Los datos y metadatos deben estar asociados con información detallada sobre su procedencia</p> <p>Los datos y metadatos deben seguir los estándares comunitarios relevantes del dominio.</p>

En los últimos años la cantidad de información recopilada ha aumentado de forma exponencial, a un nivel impensable antiguamente. Esto ha llevado a crear almacenes públicos que se centran en un área específica. Algunos de ellos pueden ser PRIDE, para la identificación de péptidos y proteínas; GEO, que almacena datos de la secuencia funcional del genoma; o GTex, que cubre más de 50 muestras de tejidos (103).

Pero un individuo no es capaz de trabajar con todo este volumen de información, es necesario recurrir a herramientas informáticas para poder analizar y comparar toda esta información y poder extraer conclusiones. Aquí es donde entra la inteligencia artificial, ya que los métodos estadísticos convencionales no tienen suficiente capacidad.

La medicina de precisión es un campo que está avanzando muy rápidamente. El uso de los SNPs de un individuo para predecir riesgos individuales de enfermedad puede hacer posible este tipo de medicina, pero para ello es necesario el tratamiento de información interactiva altamente dimensional. El aprendizaje automático es un eficaz sistema capaz de clasificar y estratificar con elevada precisión riesgos individuales de diferentes tipos de enfermedad en base a los resultados del genotipo genético (106).

1.7.2.- Machine Learning (árboles de decisión)

En la conferencia de Dartmouth, en 1956, es donde se considera que se fundó la inteligencia artificial como una disciplina académica, definida en términos generales como la inteligencia demostrada por las máquinas, en contraste a la inteligencia natural presente en los humanos y otros animales (107).

La inteligencia artificial se puede clasificar en tres grupos: analítica, inspirada en humanos o humanizada. La inteligencia artificial analítica presenta únicamente características relacionadas con la inteligencia cognitiva, es capaz de hacer predicciones basándose en el aprendizaje de experiencias anteriores. La inteligencia artificial inspirada en humanos presenta una inteligencia emocional y comprensión además de la parte cognitiva. Finalmente, la inteligencia artificial humanizada son sistemas capaces de desarrollar inteligencia social y emocional además de la cognitiva y tienen autoconciencia y conciencia de sí mismos al interactuar con otros. La inteligencia artificial tiene múltiples aplicaciones, algunas de ellas pueden ser los sistemas de reconocimiento de voz, visión computacional, bioseguridad, sistemas de control de robots, acelerar las ciencias empíricas, etc. (107).

El aprendizaje automático (Machine Learning), que entraría dentro de la inteligencia artificial analítica, se basa en una mezcla de ciencia computacional y estadística. El aprendizaje automático se puede clasificar en supervisado, semi-supervisado o no supervisado. Cuando se encuentran patrones en información sin etiquetar se habla de no supervisado, el supervisado usa datos etiquetados para inferir la clasificación, y el semi-supervisado se lleva a cabo con una pequeña cantidad de datos etiquetados y una gran cantidad de datos sin etiquetar. En concreto, en este estudio será este último tipo el que se utilice, ya que a partir de unas bases de datos introducidas inicialmente y mediante unos algoritmos previamente establecidos, se llevará a cabo un entrenamiento y a través de un posterior aprendizaje autónomo, se lograrán optimizar las agrupaciones de los datos según se van ampliando los datos recogidos. Esto permite una posible predicción al poder comparar eficazmente los nuevos datos con los que había inicialmente (107,108).

Los árboles de decisión son estructuras de decisión, que subdividen los datos recurrentemente para clasificar cada muestra. El árbol de decisión podría definirse como un **algoritmo predictivo supervisado de aprendizaje automático que divide el espacio de los predictores agrupando observaciones con valores similares para la variable respuesta o dependiente**. La realización de un árbol de decisión requiere dividir el espacio muestral en sub-regiones aplicando una serie de reglas o decisiones, con el fin de que cada sub-región contenga la mayor proporción posible de individuos de una de

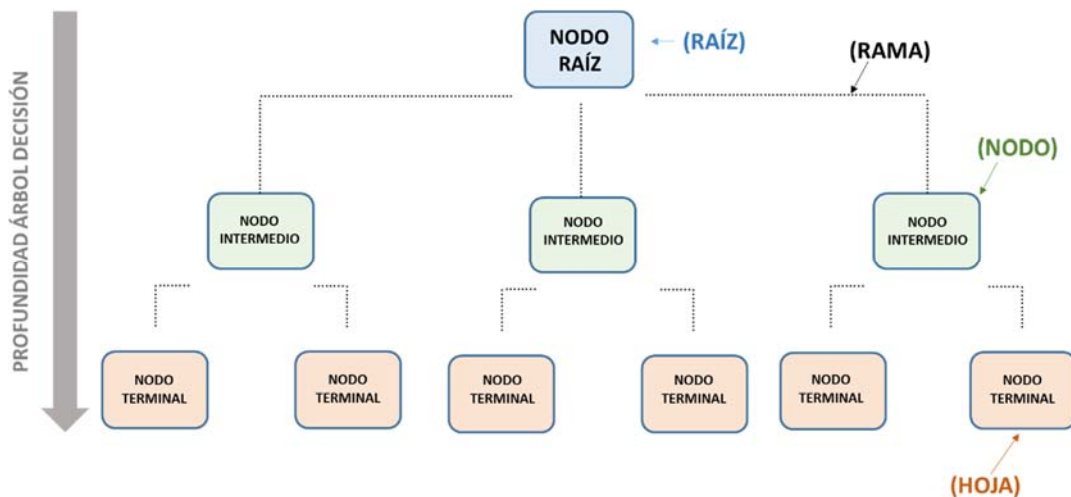
las poblaciones o categorías analizadas. Si una sub-región contiene datos de diferentes clases, se subdivide en regiones más pequeñas hasta fragmentar el espacio en sub-regiones menores que integran datos de la misma clase (108,109).

La lectura del árbol de decisión se realiza de arriba hacia abajo y tiene una estructura básica muy bien definida, formado por distintos tipos de nodos (véase figura 5):

- **Primer nodo o nodo raíz:** en él se produce la primera división en función de la variable más importante.
- **Nodos internos o intermedios:** tras la primera división encontramos estos nodos, que vuelven a dividir el conjunto de datos en función de las variables.
- **Nodos terminales u hojas:** se ubican en la parte inferior del esquema y su función es indicar la clasificación definitiva.

Cuanto mayor sea el número de nodos de una rama, el árbol de decisión tendrá mayor profundidad.

Figura 5. Esquema creación árbol de decisión.



Una característica interesante de los árboles de decisión es la facilidad con la que se puede derivar, a partir de un árbol de decisión, un conjunto de reglas de producción (del tipo IF-THEN) completamente equivalente al árbol original. Este hecho es fundamental para facilitar la comprensión del modelo de clasificación construido cuando el árbol de decisión es complejo, ya que conforme el tamaño los árboles de decisión aumenta, su inteligibilidad disminuye.

En el campo de la biomedicina, la Universidad de Stanford desarrolló en 1970 un sistema experto para ayudar al diagnóstico clínico en trastornos de la coagulación e infecciones. El sistema se llamó MYCIN y se desarrolló para identificar bacterias que causaban infecciones severas, como la meningitis. Además, mediante el sistema MYCIN se podían recomendar antibióticos dosificados, basándose en el peso del paciente. Por eso, el nombre de los programas basados en antibióticos tiene muchas veces el sufijo “mycin”. El sistema MYCIN ofreció un tratamiento correcto de los pacientes en un 65% de los casos, porcentaje muy superior al obtenido por cinco expertos, que alcanzaron un tratamiento correcto en una franja muy amplia, de entre el 42,5% y el 62%, siempre inferior, en todo caso, a la del sistema MYCIN. Sin embargo, y a pesar de estos prometedores resultados, MYCIN nunca llegó a utilizarse en la práctica debido a problemas administrativo-legales, que se planteó la cuestión de quién era el responsable del diagnóstico si se diese el caso de que el programa ofreciese un diagnóstico erróneo, con consecuencias negativas para el paciente o incluso su muerte.

Ya han transcurrido varias décadas desde el desarrollo de MYCIN, desarrollándose desde entonces otras técnicas de inteligencia artificial como los algoritmos predictivos a través del desarrollo de árboles de decisión. A diferencia de otras formas de Machine-learning, los árboles de decisión tienen la ventaja de que su recorrido puede seguirse retrospectivamente para conocer el camino seguido hasta la obtención del resultado que sea al caso. Esto permite conocer cómo se ha llegado hasta las diferentes decisiones dentro del árbol. En el mundo de la biomedicina, y particularmente para la predicción, diagnóstico y decisión de tratamientos, muchas veces es de gran importancia conocer la trazabilidad de la decisión.

Las nuevas tecnologías de análisis de datos, como los árboles de decisión basados en el aprendizaje automático, han sido identificados como herramientas potencialmente útiles para la predicción de enfermedades. En este sentido, esta tecnología llegó a valorarse como una aplicación potencial en el diseño de recomendaciones para la prevención precoz del sobrepeso y la obesidad, en cambios en la dieta y factores relacionados con el estilo de vida entre otros (110).

2. HIPÓTESIS

El conjunto de herramientas predictivas posibilita anticipar más y con mayor nivel de confianza el riesgo de padecer una enfermedad, posibilitando la opción de tomar medidas preventivas específicas frente a la aparición de la enfermedad. La aportación a la medicina predictiva de la ciencia de los datos mediante inteligencia artificial está siendo incorporada progresivamente al conjunto de elementos de predicción del que dispone el profesional sanitario para la elección de tratamientos, diagnósticos y pronósticos. Diferentes estudios han reportado la asociación entre diferentes variantes genéticas y sobrepeso/obesidad, por lo que podrían ser utilizados desde un abordaje predictivo. Por todo ello, sería plausible aplicar y desarrollar árboles de decisión (técnica de machine learning) para analizar de forma conjunta variables no modificables como el género, edad y variantes genéticas de alta frecuencia en la población (polimorfismo), identificando patrones de predicción de riesgo de sobrepeso/obesidad y comorbilidades asociadas.

3. OBJETIVOS

El objetivo principal del presente trabajo doctoral es desarrollar un análisis piloto basado en la generación de árboles de decisión para predecir el efecto de las interacciones de género, edad y diferentes variantes genéticas (single nucleotide polymorphisms, SNP) en genes implicados en diferentes aspectos nutricionales para predecir el riesgo de desarrollar sobrepeso/obesidad.

Para ello se marcarán los siguientes objetivos específicos:

1. Genotipar diferentes variantes genéticas (polimorfismo de un único nucleótido) en genes implicados en el ciclo saciedad-hambre, metabolismo de carbohidratos y grasas y consumo calórico durante el ejercicio físico.
2. Valorar mediante estudios estadísticos convencionales si el análisis genético de polimorfismos implicados en el ciclo saciedad-hambre, metabolismo de carbohidratos y grasas y consumo calórico durante el ejercicio físico se asocian con el riesgo de sobrepeso/obesidad.
3. Determinar si el análisis basado en la generación de árboles de decisiones permite predecir el riesgo de sobrepeso/obesidad según el perfil genético de los polimorfismos implicados en el ciclo saciedad-hambre, metabolismo de carbohidratos y grasas y consumo calórico durante el ejercicio físico.
4. Determinar si el análisis basado en la generación de árboles de decisiones permite predecir el riesgo de sobrepeso/obesidad en base a la interacción entre las variables género, edad y resultados del análisis genético de los polimorfismos implicados en el ciclo saciedad-hambre, metabolismo de carbohidratos y grasas y consumo calórico durante el ejercicio físico.

4. MATERIAL Y MÉTODOS

4.1 .- POBLACIÓN DE ESTUDIO

En el estudio se incluyeron 151 individuos, seleccionados de forma aleatoria de una cohorte anónima de 1500 individuos pertenecientes a población rural demográficamente estable de Castilla La Mancha (Factores de riesgo cardiovascular en una población rural de Castilla-La Mancha" (111). Esta cohorte se completó en un período de 2 años (1994-1996) formada por individuos mayores de 24 años, con residencia en el Área Sanitaria de Talavera de la Reina (Toledo). El muestreo de la cohorte se realizó sobre tamaño muestral de más de 90.000 habitantes, distribuidos en 77 municipios con una población inferior a 5.000 habitantes/municipios y actividades económicas predominantemente agrícolas y ganaderas.

El método de muestreo se realizó en dos etapas. En una primera etapa se seleccionaron aleatoriamente 16 municipios y posteriormente también aleatoriamente se seleccionaron los 1.500 sujetos de la cohorte. El número de sujetos seleccionados en cada municipio fue proporcional a su tamaño. Con la autorización del ayuntamiento correspondiente para acceder al padrón municipal se realizó la selección aleatoria de los sujetos de estudio. Esta selección se realizó por estratos de edad y género, utilizando los padrones municipales como base de muestreo.

Tras una entrevista con el equipo sanitario de la zona, se invitó a participar a los sujetos mediante una carta personal en la que se solicitó su consentimiento por escrito, explicándoles los objetivos del estudio, el tipo de entrevista y las determinaciones antropométricas. Los sujetos que no acudían a la cita eran contactados de nuevo bien con una visita a su domicilio o bien telefónicamente. La investigación fue aprobada por Comité Local de Ética del Hospital Clínico San Carlos (protocolo # C.I. 14/442-E).

Todos los individuos cumplieron los siguientes criterios de selección:

- a) Edad superior a 24 años
- b) Más de 6 meses de residencia en la región seleccionada desde el reclutamiento
- c) No presentar patologías graves o agudas
- d) No estar hospitalizado
- e) Estar en plenas facultades mentales para cumplimentar diferentes cuestionarios

- f) No estar embarazada
- g) Firmar el consentimiento informado

4.2 .- MEDIDA ÍNDICE DE MASA CORPORAL

Dos enfermeros adiestrados realizaron las extracciones de sangre para el posterior análisis genético y determinaron el peso en kg y la estatura en cm de los 151 individuos incluidos en el estudio. Se utilizaron básculas y tallímetros calibrados, los participantes fueron pesados y tallados con ropa ligera, sin zapatos y a primera hora de la mañana. El índice de masa corporal (IMC) se calculó mediante la fórmula: $\text{peso}/\text{estatura}^2$ (kg/m^2). Los participantes se clasificaron en los grupos de normopeso (IMC comprendido entre 18.50-24.99 kg/m^2) y sobrepeso-obesidad (IMC mayor o igual a 25 kg/m^2) acorde a la clasificación establecida por la Organización Mundial de la Salud.

Adicionalmente también se recogieron la edad y el género de los participantes.

4.3 .- ANÁLISIS GENÉTICO

a) Aislamiento y cuantificación de ADN

Para la extracción de los núcleos celulares se realizó un tratamiento de la sangre con tampón STMT (Tris-HCl, Tritón, sacarosa), seguido de una centrifugación 4000 rpm, 20 min a 4°C, con el objetivo de recoger los núcleos celulares sedimentados. Los núcleos fueron tratados con un tampón hipertónico NaCl 6 mol/L para la ruptura de la membrana nuclear y con proteasa K (20 mg/mL) a 37°C. Posteriormente, se realizó una centrifugación a 3500 rpm durante 20 minutos a temperatura ambiente, recogándose el DNA solubilizado del paciente en el sobrenadante. Para la recogida del DNA se añadió 1 volumen de isopropanol (1vol:1vol). Cuando precipitó la hebra de DNA se hidrató para su posterior cuantificación y amplificación por técnica PCR (polymerase chain reaction) (112,113).

La cuantificación del DNA se realizó mediante un espectrofotómetro en las longitudes de onda 260 nm y 280 nm para detectar una posible contaminación por proteínas y a 320 nm para detectar una posible contaminación por hidratos de carbono. Se calculó el ratio $\text{Abs}(260\text{nm})/\text{Abs}(280\text{nm})$ para cuantificar la pureza del ADN. Todas la muestras

dieron valores comprendidos entre 1.8-2.0, descartándose contaminación por proteínas. Teniendo en cuenta que 1 D.O = 50µg/ml de DNA bicatenario. Se realizó una solución con una concentración 10ng/µl (dilución de trabajo) (112,113).

b) Análisis de genotipado

Se seleccionaron seis polimorfismos de un único nucleótido en seis genes que habían sido potencialmente asociados con el valor del IMC y/o metabolismo y asimilación de nutrientes (21,22,88,114-118). En la tabla 6 se muestran los polimorfismos seleccionados para el presente estudio.

Tabla 6. Relación de polimorfismos que se usan en el estudio.

Gen	Identificador alteración genética	Sustitución de nucleótidos	Efecto modulador del polimorfismo
UCP2	rs660339	A>G	Termogénesis mitocondrial
AgRP	rs5030980	T>C	Control de ingesta
PPARG	rs1801282	G>C	Metabolismo de carbohidratos y lípidos
ADRB3	rs4994	A>G	Regulación lipólisis
ADRB2	rs1042714	C>G	Regulación lipólisis
INSIG2	rs7566605	C>G	Regulación adipogénesis

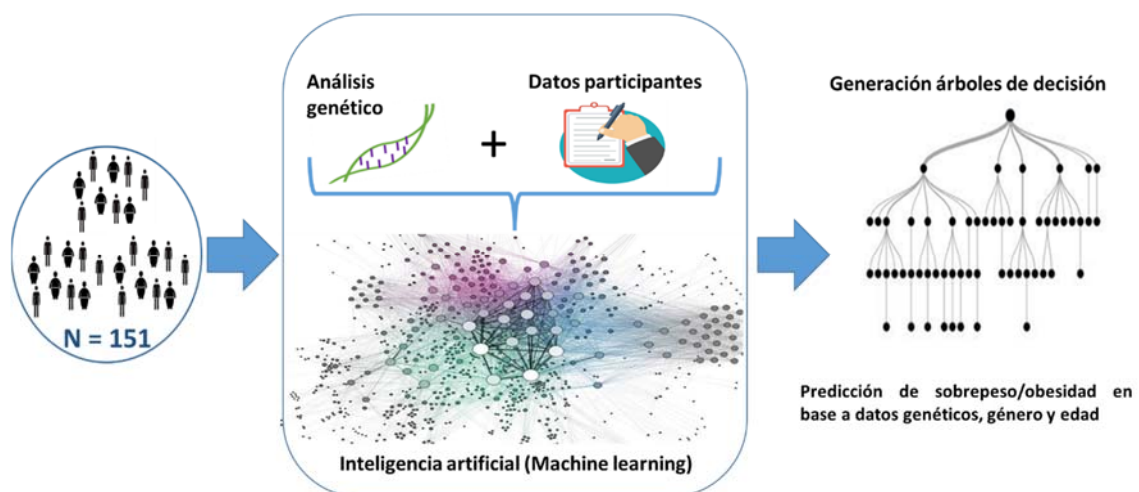
Los polimorfismos seleccionados se genotiparon mediante la técnica de PCR cuantitativa a tiempo real, utilizando sondas TaqMan. De forma breve, los ensayos de identificación de genotipos TaqMan se componen de pares de primers de PCR preoptimizados y de dos sondas fluorescentes para la discriminación alélica. Los ensayos de identificación de genotipos TaqMan se utilizan para amplificar y detectar determinados alelos en ADN genómico (ADNg). (<https://www.thermofisher.com/es/es/home/life-science/pcr/real-time-pcr/real-time-pcr-learning-center/real-time-pcr-basics/how-taqman-assays-work.html>).

Cada ensayo de genotipado contenía:

- Un par de primers sin etiquetar
- Dos sondas TaqMan (una con una etiqueta de colorante FAM y otra con una etiqueta de colorante VIC) en el extremo 5' y uniones al surco menor (MGB) y supresores no fluorescente (NFQ) en el extremo 3'.

Los cebadores TaqMan se obtuvieron mediante la aplicación TaqMan™ Gene Expression (Thermofisher.com). Las mezclas de reacción de 10 ul contenían 1 ng de ADNg como plantilla de PCR, 1 x TaqMan Gene Expression PCR Master Mix (Ref. 4331182, Applied Biosystem. Thermofisher) y 0,6 x Genotyping TaqMan-probe test (Ref. 4351374, Applied Biosystem. Thermofisher). Las placas de PCR en tiempo real se procesaron en un sistema ABI fast-7900HT (Applied Biosystems, Thermo Fisher) en condiciones de funcionamiento estándar (95°C durante 10 min y 40 ciclos de dos pasos que constan de 95 °C durante 15 segundos y 60 °C durante 1 minuto). Los genotipos se diferenciaron analizando los niveles de fluorescencia de los productos de PCR utilizando el software Sequence Detection System 2.4 (Applied Biosystems). Los genotipados se realizaron a ciegas en los servicios centrales del Parque Científico de Madrid, en Cantoblanco (fpcm.es).

Figura 6. Esquema diseño de estudio.



4.4.- GENERACIÓN DE ÁRBOLES DE DECISIÓN

Los árboles de decisión generados se representan en forma de gráficos. Estos gráficos simulan las diferentes alternativas posibles cuando se plantea una pregunta con el objetivo de determinar la secuencia óptima de decisiones que se deben realizar. A medida que el algoritmo avanza el árbol va creciendo y los subconjuntos son menos numerosos, partiendo de lo general a lo específico, a partir de la generación de los árboles de decisión, se puede predecir un resultado o riesgo en función de las características de los nodos o sujetos incluidos.

Para la realización de los árboles de decisión, se utilizó el algoritmo C4.5 ampliamente utilizado, ya que se considera el modelo estándar en la clasificación supervisada y de toma de decisiones. C4.5 es un híbrido entre abordajes CART y C4 que permite usar como regla de división la ganancia de información, el índice de diversidad de Gini o el criterio de proporción de ganancia. Además, incluye la posibilidad de realizar una post-poda pesimista del árbol. Para cada atributo discreto, se estimó una prueba con n resultados, donde n era el número de posibles valores que podía tomar un atributo o variable (género, IMC, presencia de alelos genéticos, etc). Una prueba binaria (1,0) se realizó para los atributos continuos. En cada nodo, el sistema debía decidir que prueba escoger entre tres posibles, para dividir los datos o ramas:

- La *prueba standard para cada variable discreta*, con un resultado y una rama para cada posible valor de la variable.
- La prueba de variable discreta, por la que los posibles valores se asignaron a un grupo.
- Si una variable X tenía valores numéricos continuos (edad), se realizó la prueba binaria con resultados $X \leq Z$ y $X > Z$, para lo cual debía determinarse de manera intrínseca el límite Z .

Para la generación del árbol se utilizaron las aplicaciones disponibles en las bibliotecas de R (<https://CRAN.R-project.org/package=rpart>), estableciéndose el límite de aprendizaje en 15, es decir, el aprendizaje se paraba cuando 15 o menos observaciones se encontraban en un nodo, para evitar la construcción de árboles demasiado complejos (119). El propósito de los árboles de decisión en esta tesis fue principalmente exploratorio, lo que quiere decir que no había división entre los datos

de entrenamiento y los datos de test, y, consecuentemente, las medidas de precisión comunes como AUC o las pérdidas binarias no fueron consideradas.

4.5 .- ANÁLISIS ESTADÍSTICO

La determinación de equilibrio de distribución de la frecuencia alélica de las variantes analizada se estimó mediante el equilibrio Hardy-Weinberg (HWE), basado en un test chi cuadrado. El resto de variables se sometió un test de normalidad Kolmogorov. Las variables continuas (edad) se compararon entre los grupos de interés mediante el test paramétrico de *t* de Student. Los datos de variables continuas se presentan como la media \pm EEM, a no ser que se indique de otra forma. Las diferencias en la distribución de los diferentes genotipos se estimaron mediante el test de chi-cuadrado, ajustando por Fisher en los casos necesarios. Las diferencias se consideraron estadísticamente significativas cuando el valor $P < 0,05$. El procesado de datos y el análisis estadístico se llevó a cabo con el programa estadístico SPSS (versión 25.0).

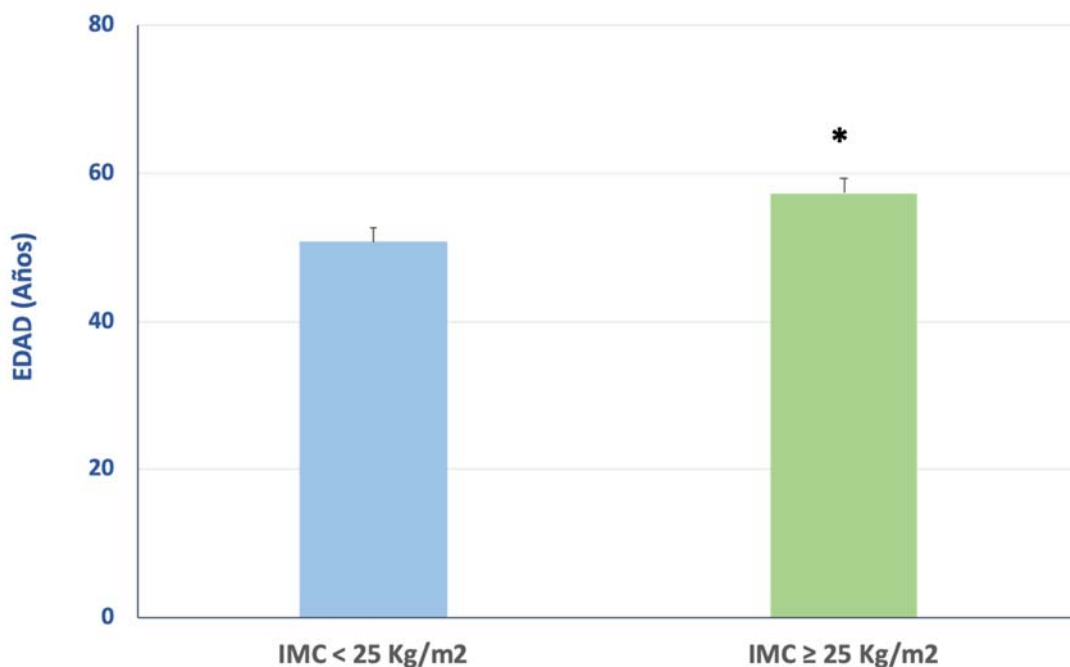
5. RESULTADOS

5.1.- DISTRIBUCIÓN POBLACIÓN ESTUDIO

Se reclutaron 151 participantes con una edad media $55,10 \pm 1,34$ años y una balanceada distribución por género, 78 hombres (51,7 %) con una edad media de 55.64 ± 1.82 años y 73 mujeres (48,3%) con edad media de $54,64 \pm 2,01$. Cuando se compararon las edades entre los dos géneros no se encontraron diferencias significativas ($p=0,35$).

A partir de la cohorte población reclutada, se generaron dos grupos comparativos en función del IMC, un grupo con normopeso (IMC entre $18,50-24,99 \text{ Kg/m}^2$) y otro de sobrepeso-obesidad (IMC $\geq 25 \text{Kg/m}^2$) con un tamaño muestral de 51 (34%) y 100 (66%), respectivamente. Como se observa en la figura 7, la edad fue levemente superior en los individuos con sobrepeso-obesidad comparado con aquellos con normopeso, alcanzando significancia estadística ($p=0,03$).

Figura 7. Comparación variable edad según índice de masa corporal.



Resultado expresado como media \pm error estándar de la media. * $p<0,05$ respecto al grupo IMC < 25 Kg/m²

5.2.- RESULTADOS DE GENOTIPADO

Como se observa en la tabla 7, las frecuencias de genotipado de cinco variantes de las genéticas analizadas eran consistentes con el equilibrio de Hardy-Weinberg (rango del valor de p de 0,132 a 0,905). Únicamente, las frecuencias de genotipado del polimorfismo analizado en el gen AgRP, codificante para la proteína relacionada agoutí, mostraron una desviación significativa del equilibrio Hardy-Weinberg, lo que se podría atribuir a un error de tipo I.

Tabla 7. Análisis equilibrio de Hardy-Weinberg de frecuencia de los polimorfismos de interés

GEN (Polimorfismo)	HWE valor P	Frecuencia alelo minoritario	Alelo minoritario	Alelo ancestral
UCP2 (<i>rs660339</i>)	0,665	0,421	A	G
AgRP (<i>rs5030980</i>)	<0,001	0,252	T	C
PPARG (<i>rs1801282</i>)	0,905	0,099	G	C
ADRB3 (<i>rs4994</i>)	0,814	0,050	G	G
ADRB2 (<i>rs1042714</i>)	0,649	0,351	G	G
INSIG2 (<i>rs7566605</i>)	0,132	0,371	C	G

Las frecuencias de genotipado para los polimorfismos en los seis genes testados en la población estudiada acorde al IMC se muestran en la Tabla 8. Los resultados de la comparación de la distribución de cada genotipo para cada polimorfismo únicamente mostraron diferencias significativas para polimorfismo rs5030980 del gen AgRP. Los datos sugieren un mayor porcentaje de individuos con IMC ≥ 25 Kg/m² portadores de la variación alélica Ala67Ala con respecto a los individuos con IMC <25 Kg/m². La variante alélica Thr67Ala en el gen AgRP (rs5030980) era más frecuente en individuos con IMC <25 Kg/m² que en los que tienen IMC ≥ 25 Kg/m². El análisis estadístico de la distribución alélica del polimorfismo (rs104714) en el gen ADRB2 entre los individuos que muestran IMC < 25 Kg/m² y aquellos con IMC ≥ 25 Kg/m² mostraron una tendencia muy cercana a

la significancia estadística, aunque no se llegó a alcanzar ésta ($p=0,05$) Tabla 8. De hecho, la presencia de la variante Gln27Gln era claramente más frecuente en la población con $IMC < 25 \text{ Kg/m}^2$ comparado con los individuos con $IMC \geq 25 \text{ Kg/m}^2$ Tabla 8. Respecto al resto de polimorfismos analizados en los genes UCP2, PPARG, ADRB3 e INSIG2 no se encontraron diferencias significativas en la frecuencia de los genotipos Tabla 8.

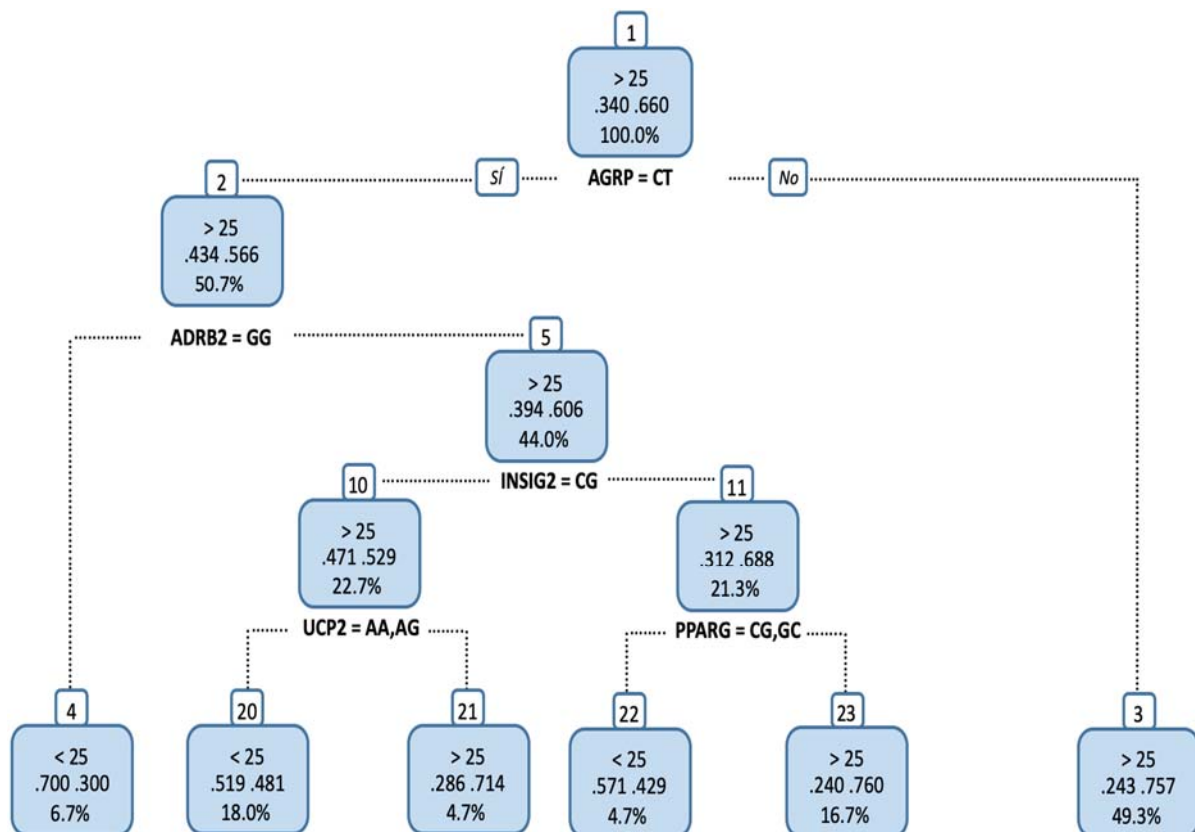
Tabla 8. Frecuencia de genotipados de los polimorfismos de interés en función del Índice de Masa Corporal (IMC).

Gen	Variante	Frecuencia de genotipo (casos / %)		Valor P
		IMC < 25 Kg/m ² (n = 51)	IMC ≥ 25 Kg/m ² (n = 100)	
UCP2 (rs660339)	AA (Val55Val)	10/19,6	14/14,0	0,66
	AG (Val55Ala)	26/51,0	53/53,0	
	GG (Ala55Ala)	15/29,4	33/33,0	
AgRP (rs5030980)	CC (Ala67Ala)	18/35,3	57/57,0	0,01
	CT (Ala67Thr)	33/64,7	43/43,0	
	TT (Thr67Thr)	0/0,0	0/0,0	
PPARG (rs1801282)	CC (Pro12Pro)	39/76,5	83/83,0	0,28
	CG (Pro12Ala)	11/21,6	17/17,0	
	GG (Ala12Ala)	1/2,0	0/0,0	
ADRB3 (rs4994)	AA (Trp64Trp)	45/88,2	91/91,0	0,59
	AG (Trp64Arg)	6/11,8	9/9,0	
	GG (Arg64Arg)	0/0,0	0/0,0	
ADRB2 (rs1042714)	CC (Gln67Gln)	23/45,1	38/38,0	0,05
	CG (Gln67Glu)	19/37,3	55/55,0	
	GG (Glu67Glu)	9/17,6	7/7,0	
INSIG2 (rs7566605)	CC (región no codificante)	4/7,8	11/11,0	0,73
	CG (región no codificante)	27/52,9	55/55,0	
	GG (región no codificante)	20/39,2	34/34,0	

5.3.- APRENDIZAJE DE LOS ÁRBOLES DE DECISIÓN ASOCIANDO EL IMC CON LOS POLIMORFISMOS GENÉTICOS

El sistema de aprendizaje dividió a los 151 individuos en dos ramas de acuerdo al polimorfismo Ala67Thr (rs5030980) del gen AgRP (Figura 8). El 50,7% de la población total incluida mostró el polimorfismo Ala67Thr (alelos CT). El 56,6% de ellos tenía un IMC ≥ 25 Kg/m² y el 43,4% mostraba un IMC < 25 Kg/m² (Figura 8). Los restantes 49.3% de la población total no mostraban ninguna variante Ala67Thr en el polimorfismo AgRP rs5030980 y el 75,7% de estos individuos estaban asociados con un IMC ≥ 25 Kg/m² (Figura 8).

Figura 8. Árbol de decisión basado en polimorfismos y el índice de masa corporal



Del nodo descendente del polimorfismo Ala67Thr rs5030980 (CT variante alélica) del gen AgRP, apareció otra nueva rama descendente (Figura 8). Esta nueva rama indicaba correspondía a la población que mostraba la variante Ala67Thr en el polimorfismo

rs5030980 del gen AgRP (el 50.7% de la población total), sólo el 6,7% de ellos mostraban la variante Gln27Gln (alelo GG) en el polimorfismo rs1042714 del gen ADRB2. De todos estos individuos, el 70% tenían un peso normal, IMC <25Kg/m². Además, de la población que mostraba la variante Ala67Thr en el polimorfismo rs5030980 del gen AgRP, el 44% de ellos también mostraban las variantes alélicas GC o CC en el polimorfismo rs1042714 del gen ADRB2 (Figura 8).

Una nueva rama derivada del nodo de esta población mostraba las variantes alélicas GC o CC en el polimorfismo rs1042714 del gen ADRB2, indicando que si el polimorfismo rs7566605 del gen INSIG2 tenía la variante alélica GC (52,9% de los individuos con variantes genéticas GC o CC en ADRB2), los individuos presentaban un IMC ≥25 Kg/m². Por otro lado, si el polimorfismo (rs660309) del gen UCP2 era GG, el 71,4% de los individuos desarrollaron un IMC ≥25 Kg/m² (Figura 8).

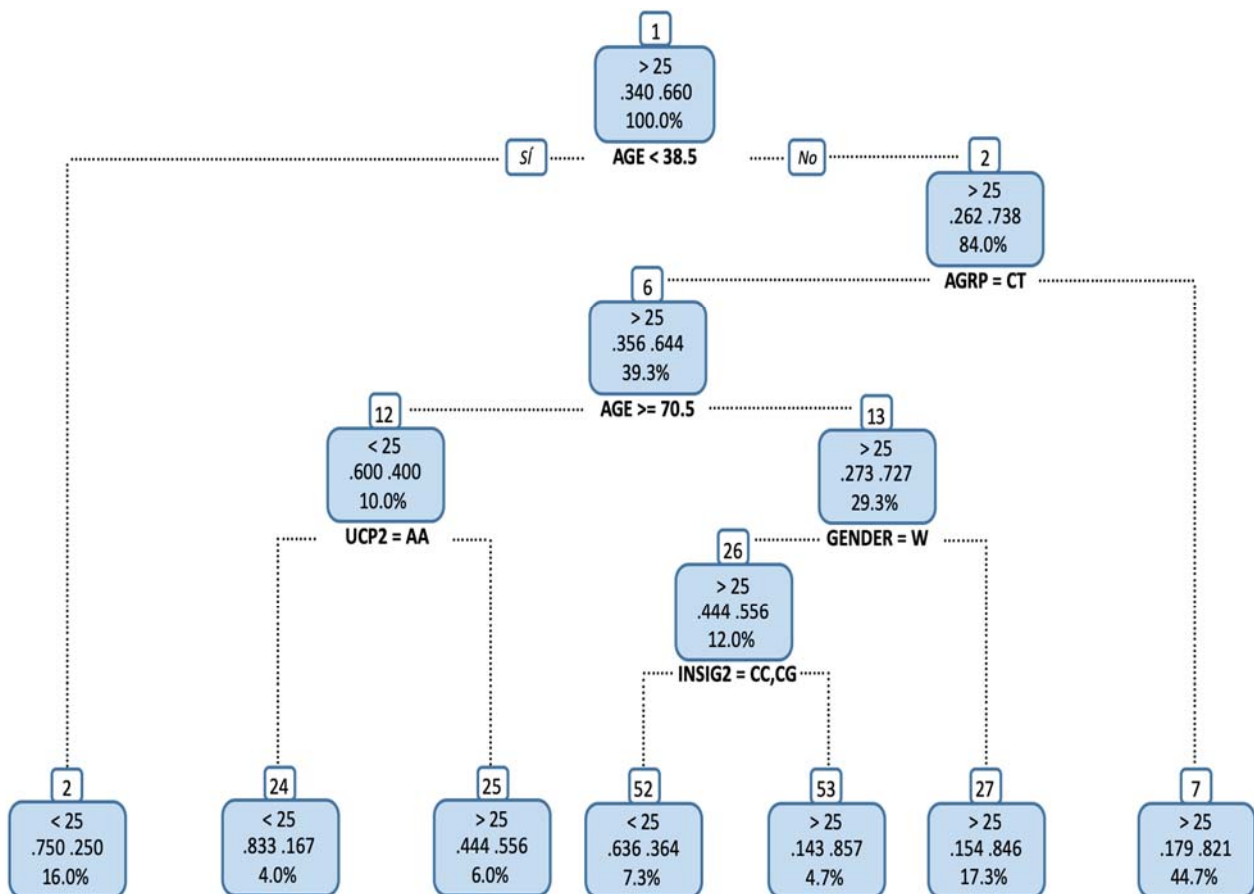
Volviendo al nodo INSIG2, del 44% de la población que mostraba la variante alélica GC o CC en el polimorfismo rs1042714 del gen ADRB2 (Figura 8), el 21,3% de ellos también mostraban las variantes alélicas CC o GG en el polimorfismo rs7566605 del gen INSIG2. En esta población, el 68,8% de los individuos tienen un IMC ≥25 Kg/m². Descendiendo por esta rama, se identificó otro nuevo nodo en el aprendizaje del árbol de decisión que sugería que si el polimorfismo rs1801282 del gen PPARG tiene la variante alélica GC (el 4,7% de la población tiene una variante diferente a la CG en el polimorfismo rs7566605 del gen INSIG2), el 42,9% de estos individuos mostrarán sobrepeso/obesidad IMC ≥ 25Kg/m². Por otro lado, si el polimorfismo del gen PPARG tenía las variantes alélicas CC o GG, el 76,0% de estos individuos mostraron sobrepeso/obesidad (Figura 8).

Con todo lo anterior, el aprendizaje y el consecuente árbol de decisión mostraron que los portadores de las variaciones de alelos AgRP (rs5030980) Ala67Ala, ADRB2 (rs1042714) Gln27Glu o Glu27Glu, INSIG2 (rs7566605) 73+9802 con genotipos CC o GG y PPARG (rs1801282) con las variaciones de alelos de Ala12Ala o Pro12Pro desarrollarán con mayor probabilidad sobrepeso/obesidad IMC≥25 Kg/m², respecto a los no portadores de las variantes descritas.

5.4.- IMPACTO DE LA EDAD Y EL GÉNERO EN EL APRENDIZAJE DEL ÁRBOL DE DECISIÓN

El aprendizaje del árbol de decisión se puede complicar, y potencialmente aumentar su precisión, añadiendo otros factores, como la edad y el género. Incluyendo estos dos factores, el sistema de aprendizaje dividía la población en dos ramas según la edad de los individuos, marcando de manera interna el punto de corte en 38,5 años, es decir, el sistema con los nodos de aprendizaje creados generó dos grupos de individuos, un grupo de edad $\geq 38,5$ años y otro de $< 38,5$ años de edad (Figura 9).

Figura 9. Árbol de decisión basado en género, edad, polimorfismos y el índice de masa corporal



Como muestra la Figura 9, sólo el 16% de la población estudiada tenía una edad inferior a 38,5 años. Como era de esperar, el 75% de estos individuos jóvenes mostraban

un IMC <25 Kg/m². Por otra parte, el 84% de la población total tenía una edad superior a 38,5 años y el 73,8% de ellos, mostraban un IMC ≥ 25 Kg/m² (Figura 9).

Del nodo descendiente de edad $\geq 38,5$ años, las diferencias genotípicas generaron un nuevo nodo. En este nodo, el 39,3% de la población mayores de 38,5 años mostraban la variante alélica CT en el polimorfismo rs5030980 del gen AgRP, y el 64,4% de ellos tenían un IMC ≥ 25 Kg/m² (Figura 9). Además, el 82% de los individuos con edad superior a 38,5 años mostraban un IMC ≥ 25 Kg/m² cuando las variantes alélicas del polimorfismo rs5030980 del gen AgRP eran diferentes a CT (Ala67Thr) (Figura 9).

Considerando dos puntos de corte etarios 38,5 y $\geq 70,5$ años, el aprendizaje del árbol de decisión añadió una nueva rama mostrando el polimorfismo rs5030980 del gen AGRP de variante alélica CT (Figura 9). El aprendizaje del árbol de decisión determinó que el 10% de los individuos que mostraban la variante alélica CT en el polimorfismo rs5030980 del gen AgRP tenían entre 38,5 y $\geq 70,5$ años y el 60% de ellos tenían un IMC <25 Kg/m² (Figura 9). Mientras que entre los individuos con un IMC <25 Kg/m², el 4% de ellos mostraban la variante alélica AA en el polimorfismo rs660339 del gen UCP2, teniendo el 83,3% de este 4% un IMC ≥ 25 Kg/m² (Figura 9)

Por otra parte, el 29,3% de los individuos que mostraban la variante alélica CT en el polimorfismo rs5030980 del gen AgRP, no estaban en el intervalo de edad entre 38,5- $\geq 70,5$ años, y el 72,7% de estos individuos mostraban un IMC ≥ 25 Kg/m² (Figura 9). Es más, el 17,3% de estos individuos eran hombres y el 84,6% de ellos tenía un IMC ≥ 25 Kg/m² (Figura 9). El 12% de los individuos con edad fuera del intervalo 38,5- $\geq 70,5$ años y que mostraban la variante alélica CT en el polimorfismo rs5030980 del gen AgRP eran mujeres. El 55,6% de estas mujeres mostraban un IMC ≥ 25 Kg/m². Es más, el 7,3% de las mujeres con edad fuera del intervalo 38,5- $\geq 70,5$ años y con la variante alélica CT en el polimorfismo rs5030980 del gen AgRP y la variante alélica CC o CG en el polimorfismo rs7566605 del gen INSIG2. Entre esta población femenina, el 63,5% de ellos mostraban un IMC <25 Kg/m² (Figura 9). El 4,7% eran portadores de la variante alélica GG en el polimorfismo rs7566605 del gen INSIG2. De ellas, el 84,6% tenían un IMC ≥ 25 Kg/m² (Figura 9).

6. DISCUSIÓN

El presente trabajo doctoral, es un estudio piloto pionero que propone la creación de árboles de decisión con el objetivo de predecir el riesgo de desarrollar sobrepeso/obesidad en base al género, edad y presencia de diferentes polimorfismos en genes relacionados con la alimentación y obesidad. Este estudio evidencia la potencial utilidad de la inteligencia artificial y, particularmente de los árboles de decisión, en la medicina preventiva e incluso predictiva, integrando los componentes genéticos, edad y género como factores de estratificación en la predicción de obesidad y, por tanto, de diferentes patologías estrechamente asociadas a la obesidad.

En el año 2011, unos investigadores norteamericanos desarrollaron un árbol de decisión sencillo pero muy interesante para valorar el riesgo de desarrollar trombosis venosa profunda. El árbol se desarrolló analizando 515 personas, analizándose 18 factores que se consideraron de forma binaria (0 si no presentaba el factor y 1 si presentaba el factor analizado). Los investigadores introdujeron en los árboles de decisión múltiples variables para poder obtener el algoritmo de riesgo. Los modelos de los árboles de decisión fueron al principio entrenados para “enseñar” al sistema que, si la persona tenía una presión sistólica de 150 mm de Hg, ésta era una medida elevada. Una vez que el sistema fue entrenado. El sistema de inteligencia artificial desarrolló un árbol de decisión optimizado, que fue capaz de revelar los factores o condicionantes o condicionantes de riesgo de desarrollar trombosis venosa profunda, de forma más precisa que los sistemas de análisis estadísticos de regresión logística. En este sentido, los árboles de decisión generados revelaron relaciones entre diferentes variables y el riesgo de trombosis venosa profunda, que no fueron identificadas mediante las técnicas de análisis epidemiológicas tradicionales. Esta sorprendente capacidad para la identificación de predictores de los árboles de decisión, incluso con mayor sensibilidad que los análisis estadísticos tradicionales fueron descritos también por otros estudios. En este sentido, un trabajo muy reciente describió las ventajas de la utilización de los árboles de decisión frente a la regresión logística para evaluar de forma integrada el efecto de la genética y diferentes factores de riesgo clásicos de cáncer de próstata (120). Los resultados encontrados por Barnholtz-Sloan et al, evidenciaron que el proceso de generación de árboles de decisión era un procedimiento iterativo, con capacidad para

repetirse e ir aprendiendo en función de las entradas, que identificaban subgrupos de riesgo mutuamente excluyentes, que compartían factores comunes asociados con el riesgo de enfermedad y no estaban limitados por supuestos de distribución que pueden violarse en las aplicaciones de datos como en los sistemas de análisis de regresión logística (120). Por ello, los autores concluyeron que los resultados de los modelos de inteligencia artificial se podían utilizar para desarrollar intervenciones y/o tratamientos individualizados, mientras que las informaciones de los modelos de regresión se debían aplicar únicamente al escenario promedio de una población (120). Por otro lado, el modelado estándar de interacciones multiplicativas o aditivas que utilizan modelos de regresión logística requieren tamaños de muestra relativamente grandes (121). Otras de las grandes ventajas de los modelos del árbol de decisión es que se pueden incluir todos los predictores potenciales en el modelo y priorizar las variables asignándoles una estructura jerárquica real asociada con el riesgo, mientras que con los modelos de regresión tradicionales, normalmente el usuario necesita tener algún conocimiento a priori sobre qué interacciones pueden ser importantes y / o qué efectos principales pueden ser estadísticamente significativos antes de ajustar interacciones de orden superior (120).

En concordancia con las fortalezas de los modelos de árbol de decisión descritos en los trabajos anteriores, los resultados presentados en este estudio evidencian que la información obtenida del análisis estadístico convencional de la distribución de sobrepeso/obesidad y las variantes genéticas analizadas era menor que la derivada del aprendizaje con árboles de decisión. Uno de los principales resultados del presente estudio fue la identificación de relaciones de diferentes polimorfismos entre sí mismas, así como entre estas variaciones genéticas y otros factores como la edad y el género para la predicción del desarrollo de sobrepeso y obesidad. Consecuentemente, aunque en algunas investigaciones biomédicas algunos trabajos han comenzado a usar árboles de decisión con biomarcadores biomédicos (122), este es uno de los primeros trabajos usando árboles de decisión para asociar información genética relacionando genes asociados con la alimentación y la predicción de sobrepeso/obesidad.

En este trabajo de tesis doctoral, el primer nodo que apareció en el aprendizaje con árboles de decisión sugirió que, en nuestra población estudiada, la ausencia del

polimorfismo Ala67Thr AgRP se asocia más frecuentemente con $IMC \geq 25$ Kg/m². El polimorfismo Ala67Thr AgRP se ha relacionado con una mayor expresión génica del gen AgRP. El gen AgRP codifica un antagonista de la melanocortina-3 y -4 y regula el control hipotalámico de los comportamientos alimenticios, a través de la unión a los receptores de la melanocortina-3 y melanocortina-4 (MC3R, MC4R) (123). La sobreexpresión de AGRP parece estimular la alimentación y produce un aumento de peso y un crecimiento excesivos (124). El gen AGRP humano se encuentra en el cromosoma 16q22 y está compuesto por cuatro exones que codifican un péptido de 132 aminoácidos que es parcialmente homólogo a la proteína Agouti. El AGRP es sintetizado y liberado por neuronas neuropéptido Y arqueadas que se sabe que están integradas funcional y anatómicamente con neuronas POMC arqueadas (125). La función de AGRP como antagonista del receptor de melanocortina endógeno se ha demostrado mediante la administración intracerebroventricular (ICV) aguda y prolongada en modelos animales y en ratones transgénicos para AGRP (123). Huszar et al, demostraron que los animales con deficiencias del MC4R desarrollaban obesidad y trastornos metabólicos, sugiriendo que la obesidad causada por Agouti o Agouti expresados de forma ubicua venía determinada en gran medida por MC4R (126). Además, como se había descrito anteriormente, AgRP también puede ser un ligando fisiológico de MC3R, lo que también ha sido asociado a obesidad inducida por Agouti en el SNC (127,128). Hay diferentes datos en la bibliografía que sugieren una relación estrecha entre los niveles de expresión de AgRP y obesidad en modelos animales (115,129,130). En este sentido, una elevada expresión de AgRP hipotalámica se ha asociado a obesidad y diabetes en ratones, mientras que los ratones transgénicos que sobreexpresaban AgRP también eran propensos a la obesidad (131). Cuando se analizan los resultados en estudios basados en grandes cohortes de sujetos humanos, y en concordancia con los resultados descritos en este trabajo, se ha confirmado que la variante Ala67Thr en rs67483042 AgRP se ha asociado con pesos corporales bajos e incluso con anorexia nerviosa (130,132).

Descendiendo en el árbol de decisión, se puede ir describiendo la implicación de más polimorfismos genéticos, cuya combinación nos permitirá definir con mayor precisión las características genéticas de un individuo con mayor o menor riesgo de sobrepeso/obesidad. De este modo, una posible combinación alcanzada desde el primer árbol de decisión fue una combinación de polimorfismos genéticos con las siguientes

variantes alélicas: AgRP Ala12Val, ADRB2 Gln27Glu o Glu27Glu, INSIG2 73+9802 el genotipo CC o GG y PPARG Ala12Ala o Pro12Pro tendrán mayor probabilidad de identificar individuos con $IMC \geq 25 \text{ Kg/m}^2$.

El aprendizaje del árbol de decisión puede mejorar su precisión según se introducen otros factores, como factores de riesgo o factores ambientales. En este trabajo, el aprendizaje del árbol de decisión se mejoró posteriormente añadiendo la edad y el género, al ser factores conocidos que pueden influir en el desarrollo del sobrepeso/obesidad.

La primera rama alcanzada de la inclusión de estos dos nuevos factores, edad y género, reveló que, en la población estudiada, los individuos con edad superior a 38,5 años presentaban más frecuentemente sobrepeso/obesidad. Entre los individuos mayores de 38,5 años la ausencia de variante alélica CT (Ala67Thr) en el polimorfismo genético AgRP rs5030980 generó un nodo para definir una nueva rama descendiente de riesgo sobrepeso/obesidad. De hecho, en la población mayor de 38,5 años, la ausencia de la variante alélica CT aumentó el riesgo de $IMC \geq 25 \text{ Kg/m}^2$. Resulta interesante, que un estudio previo sugirió que el polimorfismo Ala67Thr estaba consistentemente asociado con una reducción del IMC en el caso de los padres, pero no en el caso de la descendencia, lo que sugiere que de alguna forma el genotipo Ala67Thr podría estar ejerciendo sus efectos de una forma dependiente de la edad (133). El polimorfismo Ala67Thr se ha asociado de forma significativa con una reducción de cada una de las cuatro variables principales de obesidad en humanos, IMC, masa grasa, porcentaje de grasa corporal y grasa visceral abdominal (133). Los resultados encontrados en el presente trabajo, junto con los descritos en otros estudios, sugieren que el polimorfismo Ala67Thr en AgRP podría proporcionar un marcador de estratificación de riesgo para desarrollar obesidad, siendo los homocigotos para Ala67Ala los de alto riesgo de obesidad y los heterocigotos Ala67Thr de protección frente a la obesidad de inicio tardío en condiciones obesigénicas.

En otro estudio, se estudió la posible asociación de Ala67Thr de AgRP con la selección de consumo de diferentes nutrientes. Los resultados mostraron en sujetos de raza blanca que los heterocigotos Ala67Thr tuvieron una proporción menor de energía total proveniente de la grasa, en comparación con individuos que eran homocigotos Ala67Ala

(Ala67Thr: 29.4% vs Ala67Ala: 31.5%, $p = 0.009$), principalmente debido a un menor consumo de grasas saturadas ($p = 0,06$) y grasas monoinsaturadas ($p = 0,01$). Por otro lado, el consumo de carbohidratos entre los heterocigotos Ala67Thr fue 2,6% unidades más altas en comparación con los homocigotos Ala67Ala (Ala67Thr: 55,1% vs Ala67Ala: 52,5%, $p = 0,03$). No se encontraron diferencias entre los diferentes genotipos Ala67Ala o Ala67Thr y la ingesta total de energía o el consumo de alcohol.

Como es de esperar en este tipo de trabajo basado en la generación de árboles de decisión, la influencia de la edad en el aprendizaje del árbol de decisión se estrechaba incluso más en una nueva rama descendente. Este nodo indicaba la predicción de sobrepeso/obesidad en individuos con edad entre 38,5-70,5 años y la variante alélica CT en el polimorfismo AgRP rs5030980. Entre ellos, el aprendizaje del árbol de decisión identificó otra variante alélica genética, AA en el polimorfismo genético UCP2 rs660339. Este nodo apunta que el 83,3% de los individuos tendrán $IMC < 25 \text{ Kg/m}^2$ si se encuentran en las siguientes condiciones: edad entre 38,5 y $\geq 70,5$ años, la variante alélica CT en el polimorfismo genético AgRP rs5030980 y la variante alélica AA en el polimorfismo genético UCP2 rs660339. Diferentes estudios han apuntado que el gen UCP2 podría estar directamente involucrado en la regulación del peso corporal, ya que su localización cromosómica tanto en ratones como en humanos coincide con una región que estudios de análisis de ligamiento han demostrado estar estrechamente asociada a la obesidad y a la diabetes tipo II (134,135). Por ello, en los últimos años se han realizado numerosos estudios sobre las variantes genéticas de la UCP2 y su contribución a la regulación del peso corporal, siendo el polimorfismo Ala55Val uno de los más importantes respecto al riesgo de obesidad (136-138). En este estudio se ha observado que la variante homocigota AA (Ala55Ala) parece disminuir el riesgo de sobrepeso/obesidad. Consecuentemente, la variante alélica Val en el polimorfismo UCP2 Ala55Val estaba asociada con un aumento de riesgo de sobrepeso/obesidad. Estos resultados son respaldados por los resultados de un meta-análisis en el que fueron analizados 12 estudios, en los que fueron reclutados 4.763 individuos (2011 casos y 2752 controles) (139). Este meta-análisis describió una frecuencia media del alelo T en el polimorfismo Ala55Val del 53,1% en los casos de sobrepeso/obesidad y del 51,7% en los controles, revelando una asociación estable del alelo T del polimorfismo Ala55Val con un mayor riesgo de sobrepeso y obesidad (139). Diferentes estudios han indicado que el

polimorfismo Ala55Val tenía efectos sobre la tasa metabólica y la actividad física nocturna (140-142). Se informó que el alelo T del polimorfismo está asociado con un menor gasto de energía y una menor oxidación de grasas (140), lo que podría estar implicado en el desarrollo del sobrepeso/obesidad y comorbilidades asociadas.

Por todo lo anterior, desde un punto estrictamente genético, la variante alélica Ala55Ala podría ser interpretada como un protector frente a sobrepeso/obesidad en la población analizada en este estudio, mientras que las variantes Val55Val y Val55Ala parecen estar asociadas al riesgo de sobrepeso/obesidad.

Diferentes estudios han implicado también a las proteínas UCPs moduladoras de la capacidad de utilización de glucosa, en la aparición de resistencia a la insulina, e incluso en la regulación de la secreción de insulina por la célula beta pancreática (143,144). Así, se ha propuesto que un aumento de la expresión de las UCPS podría aportar una defensa frente a la obesidad y la consiguiente diabetes que con frecuencia la acompaña (145-147). Es más, la variante alélica 55Val/Val en el polimorfismo UCP2 Ala55Val estaba asociada con una mayor incidencia de diabetes mellitus tipo 2 en caucásicos y afroamericanos (148). Esta observación también se publicó por otros autores (148,149). Asimismo, también se ha descrito que los individuos portadores del genotipo Val/Val del polimorfismo UCP2 Ala55Val parecen tener un menor grado de desacoplamiento de la membrana interna mitocondrial, menor gasto energético y una mayor susceptibilidad a la obesidad y a la diabetes mellitus tipo 2 (150,151). No obstante, otros estudios han publicado que este polimorfismo no está asociado individualmente con la ratio metabólico basal, el síndrome metabólico, IMC, obesidad, secreción de insulina o diabetes tipo 2 (152,138). Las inconsistencias en estos resultados publicados sobre polimorfismos se pueden explicar, al menos parcialmente, por diferentes factores que podrían introducir ruido en la comparación de los distintos estudios, como el tamaño muestral, la raza, la edad y el sexo. Como fortaleza del presente estudio, el aprendizaje de los árboles de decisión identificó características etarias muy específicas y asociaciones significativas entre los polimorfismos genéticos del gen AgRP y los portadores del genotipo Ala/Ala en el UCP2 Ala55Val con el IMC.

El tamaño, forma y distribución de los porcentajes de la composición corporal de un adulto varía sustancialmente con las variables edad y género. Diferentes trabajos han

demostrado diferencias significativas en el almacenamiento y metabolización de grasa dependientes de género (153). En comparación con los hombres, las mujeres tienen más tejido adiposo subcutáneo y triacilglicéridos intramusculares. Además, hay una mayor actividad de la lipoproteína lipasa en los adipocitos de los hombres y esto se correlaciona con las concentraciones de estrógeno (153). Durante el ejercicio aeróbico de resistencia hay una mayor oxidación de las grasas en las mujeres en comparación con los hombres, y esto puede conducir a un ahorro de glucógeno y proteínas (153). Esto puede estar relacionado con una mayor sensibilidad a la insulina en las mujeres que media un aumento en la sintasa de ácidos grasos libres y también puede cambiar con los genotipos INSIG2. Por esta razón, es importante establecer e identificar los polimorfismos genéticos que pueden estar asociados de forma diferente con el sobrepeso/obesidad según las diferencias sexuales. En este sentido, se ha descrito que en el gen INSIG2 (de sus siglas en inglés: Insulin-induced gene 2), estrechamente implicado con el metabolismo de las grasas, existe un polimorfismo que en el caso de las mujeres podría favorecer que hubiera menor acúmulo de grasa subcutánea mientras que en los hombres provoca un aumento de la acumulación de la grasa subcutánea acumulada, incluso a pesar de que los hombres realizaran ejercicio físico de tipo aeróbico. Este hecho refleja la existencia de factores que también son importantes para comprender el efecto que una variante genética puede tener. Hay muchos factores que van a modular el nivel de efecto de las alteraciones genéticas, entre ellos cabe destacar principalmente el género, la edad e incluso la raza.

En el año 2006, a través de un estudio de barrido genómico se genotiparon más de 115.000 polimorfismos de único nucleótido en 1071 participantes del Estudio Framingham, identificándose que el polimorfismo rs7566605 del gen INSIG2 se asoció significativamente con el riesgo de desarrollar obesidad (95). En este mismo estudio, se realizó el genotipado de otras poblaciones, obteniéndose que los resultados se replicaron en cuatro cohortes, incluyendo población americana, europea y afroamericana (95). Los investigadores descifraron un modelo recesivo, de tal manera, que el incremento del riesgo de obesidad tan sólo se presentaría en aproximadamente un 10% de la población. Sin embargo, estudios posteriores realizados en cohortes con importantes tamaños muestrales no fueron capaces de replicar los hallazgos iniciales. En uno de estos estudios posteriores, la variante rs7566605 fue genotipada en dos

grandes cohortes, una con 4.916 individuos provenientes del estudio europeo de investigación prospectiva del cáncer (EPIC) Norfolk (EPIC5000), que mostró un valor medio de IMC de $26,0 \pm 3,7 \text{ kg/m}^2$ y un porcentaje de obesidad del 12% (154), y en la cohorte Medical Research Council (MRC) compuesta por 1.683 individuos, con un valor medio de IMC de $27,2 \pm 4,8 \text{ kg/m}^2$ y una prevalencia de obesidad del 22% (155). Todos los individuos eran de origen caucásico y no estaban relacionados entre sí. Los resultados revelaron que la frecuencia del alelo menor (alelo C) era del 32% en ambas cohortes, no encontrándose evidencias de un mayor riesgo de obesidad en homocigotos CC en comparación con los homocigotos GG [GG: 26.1 ± 0.08 vs Portadores del alelo C: $25,8 \pm 0,07 \text{ kg/m}^2$].

Una de las principales fortalezas de la inteligencia artificial es la generación de asociaciones dinámicas entre todas las variables de estudio, según vaya optimizándose el aprendizaje del sistema. En este sentido, durante el proceso de aprendizaje el árbol de decisión reveló que entre el género femenino existía un porcentaje de mujeres portadoras de las variantes alélicas CT en el polimorfismo AgRP rs5030980 y GG en el polimorfismo genético INSIG2 rs7566605 y cuya edad no estaba incluida en los puntos de corte etarios $>38,5$ y $\geq 70,5$ años, tendrían mayor riesgo de $\text{IMC} \geq 25 \text{ Kg/m}^2$. Curiosamente, de forma diferente al resultado obtenido del árbol de decisión, un estudio sugirió que las mujeres con una copia del alelo C tenían mayores niveles de grasa subcutánea básica pero los hombres no mostraban tal asociación (156). Otros estudios publicados en poblaciones japonesas sugirieron una falta de asociación entre el SNP rs7566605 en el gen INSIG2 y tener sobrepeso (en población de mediana edad y ancianos) (157). Como se ha mencionado anteriormente, es probable que factores como la raza, la edad y asociaciones con otros SNPs puedan tener mucha influencia para determinar el impacto de SNPs específicos en el desarrollo de sobrepeso/obesidad.

Dina y colaboradores del grupo francés *Données Epidémiologiques sur le Syndrome d'Insulino-Résistance (DESIR)* también estudió a sujetos mayores resistentes a la insulina y no encontró una asociación del SNP rs7566605 con el IMC (158). El análisis de los participantes alemanes del estudio SHIP de Roskopf et al. tampoco encontraron una asociación positiva entre el SNP rs7566605 y el IMC, cuando las muestras se estratificaron por genotipo y género (159). Sin embargo, el subgrupo de individuos

obesos ($IMC > 30 \text{ kg/m}^2$) mostraron una mayor frecuencia del alelo C que las personas con sobrepeso ($IMC: 25-30 \text{ kg/m}^2$) (159). Curiosamente, un estudio reciente examinó el efecto del genotipo en la pérdida de peso (intervención de un año) y encontró que los niños con el genotipo CC perdieron menos peso que los niños con el genotipo GC / GG (160). La falta de replicación de la asociación entre el polimorfismo rs7566605 y el IMC podría deberse a limitaciones en la creación de los grupos comparativos, con baja incidencia de obesidad en muchas de las poblaciones examinadas.

Por todo lo anterior, aunque evidentemente se pueden establecer diferentes limitaciones en este estudio, como el limitado tamaño de la muestra ($n=151$), este trabajo debe ser considerado como un estudio piloto, que ha demostrado el amplio campo de aplicación de nuevas tecnologías de análisis de datos, como el aprendizaje con árboles de decisión, como herramientas útiles para la predicción de enfermedades. A este respecto, si pudiéramos predecir si un individuo tuviera mayor susceptibilidad para sufrir sobrepeso/obesidad de acuerdo con su genotipo, edad y género, podría resultar de gran aplicación en el diseño de recomendaciones preventivas precoces en cambios en la dieta y factores relacionados con el estilo de vida.

En definitiva, parece claro que siempre que se tenga que tomar una decisión condicionada, los árboles de decisión son una herramienta muy útil en biomedicina, ya que permiten ver de una manera gráfica las opciones, probabilidades y resultados que se han de tener en cuenta a la hora de predecir y prevenir. Además, los árboles de decisión son fáciles de entender y permiten visualizar gráficamente las diferentes opciones a la hora de tomar una decisión. Es evidente, por tanto, que la predicción de eventos patológicos futuros como el sobrepeso y la obesidad es uno de los campos del conocimiento que más se beneficiarán del aprendizaje dirigido, en el que disciplinas como la medicina, la biología, la física y la computación deben interactuar.

7. CONCLUSIONES

Las principales conclusiones del trabajo se enumeran a continuación:

1. Se encontraron diferentes polimorfismos de un solo nucleótido con mayor frecuencia en pacientes con sobrepeso / obesidad
2. La técnica de inteligencia artificial de learning machine, basada en la generación de árboles de decisión, identificó relaciones entre las variables genéticas y riesgo de obesidad que no fueron identificadas mediante las técnicas de análisis estadísticos tradicionales
3. La variante alélica en homocigosis Ala55Ala del gen UCP2 podría ser interpretada como un protector frente a sobrepeso/obesidad en la población analizada en este estudio, mientras que las variantes Val55Val y Val55Ala incrementan el riesgo de sobrepeso/obesidad.
4. El aprendizaje del árbol de decisión identificó un haplotipo específico asociado con alto riesgo de sobrepeso / obesidad. Los portadores de las variantes alélicas AgRP (rs5030980) Ala67Ala, ADRB2 (rs1042714) Gln27Glu o Glu27Glu, INSIG2 (rs7566605) 73+9802 con genotipos CC o GG y PPARG (rs1801282) con las variaciones de alelos de Ala12Ala o Pro12Pro desarrollan con mayor probabilidad sobrepeso/obesidad ($IMC \geq 25 \text{ Kg/m}^2$), respecto a los no portadores de las variantes anteriores.
5. La creación de árboles de decisión permite predecir el riesgo de desarrollar sobrepeso / obesidad en base a las características genéticas, género y edad.
6. Los individuos mayores de 38,5 años con ausencia de variante alélica CT (Ala67Thr) en el polimorfismo genético AgRP rs5030980 generó un nodo para definir una nueva rama descendiente de riesgo sobrepeso/obesidad. De hecho, en la población mayor de 38,5 años, la ausencia de la variante alélica CT aumentó el riesgo de $IMC \geq 25 \text{ Kg/m}^2$
7. En el género femenino, la presencia de las variantes alélicas CT en el polimorfismo AgRP rs5030980 y GG en el polimorfismo genético INSIG2 rs7566605 y una edad no incluidas en los cortes etarios $>38,5$ y $\geq 70,5$ años, tienen mayor riesgo de sobrepeso/obesidad que las mujeres no portadoras de las anteriores variantes descritos y edades inferiores a 38,5 años.
8. Este estudio respalda el amplio campo de aplicación de las nuevas tecnologías de análisis de datos.

8. BIBLIOGRAFÍA

- 1.- Poirier P, Giles TD, Bray GA, et al. Obesity and cardiovascular disease: pathophysiology, evaluation, and effect of weight loss: an update of the 1997 American Heart Association Scientific Statement on Obesity and Heart Disease from the Obesity Committee of the Council on Nutrition, Physical Activity, and Metabolism. *Circulation*. 2006;113(6):898-918. doi:10.1161/CIRCULATIONAHA.106.171016
- 2.- Avgerinos KI, Spyrou N, Mantzoros CS, Dalamaga M. Obesity and cancer risk: Emerging biological mechanisms and perspectives. *Metabolism*. 2019;92:121-135. doi:10.1016/j.metabol.2018.11.001
- 3.- Murugan AT, Sharma G. Obesity and respiratory diseases. *Chron Respir Dis*. 2008;5(4):233-242. doi:10.1177/1479972308096978
- 4.- Galbete Ciáurriz, C. Influencia de diversas variantes genéticas y el estilo de vida sobre el riesgo de obesidad en población mayor de 55 años de la cohorte SUN, Tesis Doctoral. Departamento de ciencias de la alimentación, fisiología y toxicología. Facultad Farmacia. Universidad de Navarra. 2012. <https://dialnet.unirioja.es/servlet/tesis?codigo=64724>
- 5.- Sydney, U.o., Factors associated with weight gain and the development of obesity. An overview of existing scientific literature. 2011, Boden Institute of Obesity, Nutrition, Exercise and Eating Disorders. <http://healthyweight.health.gov.au/wps/wcm/connect/healthyweight/0f878e92-9028-40bc-b3fe-73766043f373/Factors+associated+with+weight+gain+and+the+development+of+obesity.pdf?MOD=AJPERES&useDefaultText=0&useDefaultDesc=0>
- 6.- Organization W.H. Obesity and overweight. [cited 2018 12/12/2018]; Available from: <https://www.who.int/news-room/fact-sheets/detail/obesity-and-overweight>.
- 7.- Tremmel M, Gerdtham UG, Nilsson PM, Saha S. Economic Burden of Obesity: A Systematic Literature Review. *Int J Environ Res Public Health*. 2017;14(4):435. doi:10.3390/ijerph14040435
- 8.- Hruby A, Hu FB. The Epidemiology of Obesity: A Big Picture. *Pharmacoeconomics*. 2015;33(7):673-689. doi:10.1007/s40273-014-0243-x
- 9.- Bateman RM, Sharpe MD, Jagger JE, et al. 36th International Symposium on Intensive Care and Emergency Medicine : Brussels, Belgium. 15-18 March 2016 [published correction appears in *Crit Care*. 2016 Oct 24;20:347]. *Crit Care*. 2016;20(Suppl 2):94. Published 2016 Apr 20. doi:10.1186/s13054-016-1208-6

- 10.- OECD. Obesity Update 2017. [16/12/2018]; Available from: <https://www.oecd.org/els/health-systems/Obesity-Update-2017.pdf>.
- 11.- Hernández Á, Zomeño MD, Décano IR, et al. Excess Weight in Spain: Current Situation, Projections for 2030, and Estimated Direct Extra Cost for the Spanish Health System. *Rev Esp Cardiol (Engl Ed)*. 2019;72(11):916-924. doi:10.1016/j.rec.2018.10.010
- 12.- Ceballos-Macías JJ, Reynaldo Pérez Negrón-Juárez R, Flores-Real JA, Vargas-Sánchez J, Ortega-Gutiérrez G, Madriz-Prado R, Hernández-Moreno A. Obesity. Pandemia of the century XXI. *Rev. sanid. mil.* 2018; 72(5-6): 332-338.
- 13.- Romieu I, Dossus L, Barquera S, et al. Energy balance and obesity: what are the main drivers?. *Cancer Causes Control*. 2017;28(3):247-258. doi:10.1007/s10552-017-0869-z
- 14.- Leung AWY, Chan RSM, Sea MMM, Woo J. An Overview of Factors Associated with Adherence to Lifestyle Modification Programs for Weight Management in Adults. *Int J Environ Res Public Health*. 2017;14(8):922. doi:10.3390/ijerph14080922
- 15.- Keith SW, Redden DT, Katzmarzyk PT, et al. Putative contributors to the secular increase in obesity: exploring the roads less traveled. *Int J Obes (Lond)*. 2006;30(11):1585-1594. doi:10.1038/sj.ijo.0803326
- 16.- Jakicic JM, Rogers RJ, Davis KK, Collins KA. Role of Physical Activity and Exercise in Treating Patients with Overweight and Obesity. *Clin Chem*. 2018;64(1):99-107. doi:10.1373/clinchem.2017.272443
- 17.- Washburn RA, Szabo AN, Lambourne K, et al. Does the method of weight loss effect long-term changes in weight, body composition or chronic disease risk factors in overweight or obese adults? A systematic review. *PLoS One*. 2014;9(10):e109849. doi:10.1371/journal.pone.0109849
- 18.- Chin SH, Kahathuduwa CN, Binks M. Physical activity and obesity: what we know and what we need to know. *Obes Rev*. 2016;17(12):1226-1244. doi:10.1111/obr.12460
- 19.- Blundell JE, Gibbons C, Caudwell P, Finlayson G, Hopkins M. Appetite control and energy balance: impact of exercise. *Obes Rev*. 2015;16 Suppl 1:67-76. doi:10.1111/obr.12257

- 20.- Donnelly JE, Smith BK. Is exercise effective for weight loss with ad libitum diet? Energy balance, compensation, and gender differences. *Exerc Sport Sci Rev.* 2005;33(4):169-174. doi:10.1097/00003677-200510000-00004
- 21.- Takeuchi S, Katoh T, Yamauchi T, Kuroda Y. ADRB3 polymorphism associated with BMI gain in Japanese men. *Exp Diabetes Res.* 2012;2012:973561. doi:10.1155/2012/973561
- 22.- Ruiz JR, Larrarte E, Margareto J, Ares R, Labayen I. Role of β_2 -adrenergic receptor polymorphisms on body weight and body composition response to energy restriction in obese women: preliminary results. *Obesity (Silver Spring).* 2011;19(1):212-215. doi:10.1038/oby.2010.130
- 23.- Ministerio de Sanidad, S.S.e.I.S.P.e.d.S. Encuesta Europea de Salud en España 2017. Resultados. Sistema de información sanitaria. Salud y estilos de vida. 2017. http://www.ine.es/dyngs/INEbase/es/operacion.htm?c=Estadistica_C&cid=1254736176783&menu=resultados&idp=1254735573175.
- 24.- Gómez-Cabello A, Vicente Rodríguez G, Vila-Maldonado S, Casajús JA, Ara I. Envejecimiento y composición corporal: la obesidad sarcopénica en España [Aging and body composition: the sarcopenic obesity in Spain]. *Nutr Hosp.* 2012;27(1):22-30. doi:10.1590/S0212-16112012000100004
- 25.- Hallam J, Boswell RG, DeVito EE, Kober H. Gender-related Differences in Food Craving and Obesity. *Yale J Biol Med.* 2016;89(2):161-173.
- 26.- Asarian L, Geary N. Sex differences in the physiology of eating. *Am J Physiol Regul Integr Comp Physiol.* 2013;305(11):R1215-R1267. doi:10.1152/ajpregu.00446.2012.
- 27.- Gómez Canales, E. and A.I. Flores Blas. Genética en el tratamiento de la obesidad: nutrigenética y nutrigenómica en la era de la medicina personalizada y preventiva, U.A.d. Barcelona, Editor. 2015.
- 28.- Kanter R, Caballero B. Global gender disparities in obesity: a review. *Adv Nutr.* 2012;3(4):491-498. Published 2012 Jul 1. doi:10.3945/an.112.002063
- 29.- Isacco L, Miles-Chan JL. Gender-specific considerations in physical activity, thermogenesis and fat oxidation: implications for obesity management. *Obes Rev.* 2018;19 Suppl 1:73-83. doi:10.1111/obr.12779

- 30.- Gambineri A, Pelusi C. Sex hormones, obesity and type 2 diabetes: is there a link?. *Endocr Connect.* 2019;8(1):R1-R9. doi:10.1530/EC-18-0450
- 31.- Shi H, Brown LM, Rahimian R. Sex/Gender Differences in Metabolism and Behavior: Influence of Sex Chromosomes and Hormones. *Int J Endocrinol.* 2015;2015:245949. doi:10.1155/2015/245949
- 32.- Yasukochi Y, Sakuma J, Takeuchi I, et al. Identification of three genetic variants as novel susceptibility loci for body mass index in a Japanese population. *Physiol Genomics.* 2018;50(3):179-189. doi:10.1152/physiolgenomics.00117.2017
- 33.- Flier JS and Maratos-Flier E. Biología de la obesidad, in Harrison. *Principios de Medicina Interna*, 18e. 2012, McGraw-Hill Education. p. 622-629.
- 34.- Andela S, Burrows TL, Baur LA, Coyle DH, Collins CE, Gow ML. Efficacy of very low-energy diet programs for weight loss: A systematic review with meta-analysis of intervention studies in children and adolescents with obesity. *Obes Rev.* 2019;20(6):871-882. doi:10.1111/obr.12830
- 35.- Rankinen T, Zuberi A, Chagnon YC, et al. The human obesity gene map: the 2005 update. *Obesity (Silver Spring).* 2006;14(4):529-644. doi:10.1038/oby.2006.71
- 36.- Adams KF, Schatzkin A, Harris TB, et al. Overweight, obesity, and mortality in a large prospective cohort of persons 50 to 71 years old. *N Engl J Med.* 2006;355(8):763-778. doi:10.1056/NEJMoa055643.
- 37.- Bray GA. Health hazards of obesity. *Endocrinol Metab Clin North Am.* 1996;25(4):907-919. doi:10.1016/s0889-8529(05)70361-3
- 38.- Banegas JR, López-García E, Gutiérrez-Fisac JL, Guallar-Castillón P, Rodríguez-Artalejo F. A simple estimate of mortality attributable to excess weight in the European Union. *Eur J Clin Nutr.* 2003;57(2):201-208. doi:10.1038/sj.ejcn.1601538
- 39.- Anderson A, Baker JL, Breda J, Byers T, Cleary MP, Colditz G et al. *IARC Handbooks of Cancer Prevention: Absence of Excess Body Fatness.* France: International Agency for Research on Cancer , 2018. 658 p. (IARC Handbooks).
- 40.- Malone, J.I. and B.C. Hansen, Does obesity cause type 2 diabetes mellitus (T2DM)? Or is it the opposite? *Pediatr Diabetes*, 2019. 20(1):5-9.

- 41.- Koliaki C, Liatis S, Kokkinos A. Obesity and cardiovascular disease: revisiting an old relationship. *Metabolism*. 2019;92:98-107. doi:10.1016/j.metabol.2018.10.011
- 42.- Ortega-Loubon C, Fernández-Molina M, Singh G, Correa R. Obesity and its cardiovascular effects. *Diabetes Metab Res Rev*. 2019;35(4):e3135. doi:10.1002/dmrr.3135
- 43.- Bastien M, Poirier P, Lemieux I, Després JP. Overview of epidemiology and contribution of obesity to cardiovascular disease. *Prog Cardiovasc Dis*. 2014;56(4):369-381. doi:10.1016/j.pcad.2013.10.016
- 44.- Parida S, Siddharth S, Sharma D. Adiponectin, Obesity, and Cancer: Clash of the Bigwigs in Health and Disease. *Int J Mol Sci*. 2019;20(10):2519. Published 2019 May 22. doi:10.3390/ijms20102519
- 45.- Tumminia A, Vinciguerra F, Parisi M, et al. Adipose Tissue, Obesity and Adiponectin: Role in Endocrine Cancer Risk. *Int J Mol Sci*. 2019;20(12):2863. Published 2019 Jun 12. doi:10.3390/ijms20122863
- 46.- López-Suárez A. Burden of cancer attributable to obesity, type 2 diabetes and associated risk factors. *Metabolism*. 2019;92:136-146. doi:10.1016/j.metabol.2018.10.013
- 47.- Lauby-Secretan B, Scoccianti C, Loomis D, et al. Body Fatness and Cancer--Viewpoint of the IARC Working Group. *N Engl J Med*. 2016;375(8):794-798. doi:10.1056/NEJMSr1606602
- 48.- Spelta F, Fratta Pasini AM, Cazzoletti L, Ferrari M. Body weight and mortality in COPD: focus on the obesity paradox. *Eat Weight Disord*. 2018;23(1):15-22. doi:10.1007/s40519-017-0456-z
- 49.- Carneiro G, Zanella MT. Obesity metabolic and hormonal disorders associated with obstructive sleep apnea and their impact on the risk of cardiovascular events. *Metabolism*. 2018;84:76-84. doi:10.1016/j.metabol.2018.03.008
- 50.- Jiang D, Wang L, Bai C, Chen O. Association between abdominal obesity and asthma: a meta-analysis. *Allergy Asthma Clin Immunol*. 2019;15:16. doi:10.1186/s13223-019-0333-6

- 51.- Tashiro H, Shore SA. Obesity and severe asthma. *Allergol Int.* 2019;68(2):135-142. doi:10.1016/j.alit.2018.10.004
- 52.- Hammoudeh S, Gadelhak W, Janahi IA. Asthma and obesity in the Middle East region: An overview. *Ann Thorac Med.* 2019;14(2):116-121. doi:10.4103/atm.ATM_115_18
- 53.- López Farré A. *Cardio Genética.* CTO Medicina SL. 2016. Pp: 1-360. ISSN 978-84-16153-91-6
- 54.- Paidós'84. Estudio epidemiológico sobre nutrición y obesidad infantil. Madrid: Jofamar,1985.
- 55.- Zhang Y, Proenca R, Maffei M, Barone M, Leopold L, Friedman JM. Positional cloning of the mouse obese gene and its human homologue [published correction appears in *Nature* 1995;374(6521):479]. *Nature.* 1994;372(6505):425-432. doi:10.1038/372425a0
- 56.- Halaas JL, Gajiwala KS, Maffei M, et al. Weight-reducing effects of the plasma protein encoded by the obese gene. *Science.* 1995;269(5223):543-546. doi:10.1126/science.7624777
- 57.- Montague CT, Farooqi IS, Whitehead JP, et al. Congenital leptin deficiency is associated with severe early-onset obesity in humans. *Nature.* 1997;387(6636):903-908. doi:10.1038/43185
- 58.- Halberg N, Schraw TD, Wang ZV, et al. Systemic fate of the adipocyte-derived factor adiponectin. *Diabetes.* 2009;58(9):1961-1970. doi:10.2337/db08-1750
- 59.- Parente DJ, Garriga C, Baskin B, et al. Neuroleptin 2 nonsense variant associated with anxiety, autism, intellectual disability, hyperphagia, and obesity. *Am J Med Genet A.* 2017;173(1):213-216. doi:10.1002/ajmg.a.37977
- 60.- Garaulet M, Smith CE, Hernández-González T, Lee YC, Ordovás JM. PPAR γ Pro12Ala interacts with fat intake for obesity and weight loss in a behavioural treatment based on the Mediterranean diet. *Mol Nutr Food Res.* 2011;55(12):1771-1779. doi:10.1002/mnfr.201100437
- 61.- Kaulfers AM, Deka R, Dolan L, Martin LJ. Association of INSIG2 polymorphism with overweight and LDL in children. *PLoS One.* 2015;10(1):e0116340. doi:10.1371/journal.pone.0116340

- 62.- Nicoletti CF, de Oliveira AP, Brochado MJ, et al. The Ala55Val and -866G>A polymorphisms of the UCP2 gene could be biomarkers for weight loss in patients who had Roux-en-Y gastric bypass. *Nutrition*. 2017;33:326-330. doi:10.1016/j.nut.2016.07.020
- 63.- Akiyama M, Okada Y, Kanai M, et al. Genome-wide association study identifies 112 new loci for body mass index in the Japanese population. *Nat Genet*. 2017;49(10):1458-1467. doi:10.1038/ng.3951
- 64.- Song QY, Song JY, Wang Y, et al. Association Study of Three Gene Polymorphisms Recently Identified by a Genome-Wide Association Study with Obesity-Related Phenotypes in Chinese Children. *Obes Facts*. 2017;10(3):179-190. doi:10.1159/000471487
- 65.- Lindgren CM, Heid IM, Randall JC, et al. Genome-wide association scan meta-analysis identifies three Loci influencing adiposity and fat distribution [published correction appears in *PLoS Genet*. 2009;5(7). doi: 10.1371/annotation/b6e8f9f6-2496-4a40-b0e3-e1d1390c1928].
- 67.- De R, Hu T, Moore JH, Gilbert-Diamond D. Characterizing gene-gene interactions in a statistical epistasis network of twelve candidate genes for obesity. *BioData Min*. 2015;8:45. Published 2015 Dec 29. doi:10.1186/s13040-015-0077-x.
- 68.- Sharma P, Dwivedi S. Nutrigenomics and Nutrigenetics: New Insight in Disease Prevention and Cure. *Indian J Clin Biochem*. 2017;32(4):371-373. doi:10.1007/s12291-017-0699-5
- 69.- Ramos-Lopez O, Milagro FI, Allayee H, et al. Guide for Current Nutrigenetic, Nutrigenomic, and Nutriepigenetic Approaches for Precision Nutrition Involving the Prevention and Management of Chronic Diseases Associated with Obesity. *J Nutrigenet Nutrigenomics*. 2017;10(1-2):43-62. doi:10.1159/000477729
- 70.- López Farré A. Plátano de Canarias y Salud. *Asprocan*. 2017.
- 71.- Kalnina I, Kapa I, Pirags V, Ignatovica V, Schiöth HB, Klovins J. Association between a rare SNP in the second intron of human Agouti related protein gene and increased BMI. *BMC Med Genet*. 2009;10:63.. doi:10.1186/1471-2350-10-63

- 72.- Li P, Tiwari HK, Lin WY, et al. Genetic association analysis of 30 genes related to obesity in a European American population. *Int J Obes (Lond)*. 2014;38(5):724-729. doi:10.1038/ijo.2013.140
- 73.- Lin WJ, Salton SR. The regulated secretory pathway and human disease: insights from gene variants and single nucleotide polymorphisms. *Front Endocrinol (Lausanne)*. 2013;4:96. Published 2013 Aug 6. doi:10.3389/fendo.2013.00096
- 74.- Yu K, Li L, Zhang L, Guo L, Wang C. Association between MC4R rs17782313 genotype and obesity: A meta-analysis. *Gene*. 2020;733:144372. doi:10.1016/j.gene.2020.144372
- 75.- Magno FCCM, Guaraná HC, da Fonseca ACP, et al. Association of the MC4R rs17782313 polymorphism with plasma ghrelin, leptin, IL6 and TNF α concentrations, food intake and eating behaviors in morbidly obese women. *Eat Weight Disord*. 2021;26(4):1079-1087. doi:10.1007/s40519-020-01003-5
- 76.- Park S, Daily JW, Zhang X, Jin HS, Lee HJ, Lee YH. Interactions with the MC4R rs17782313 variant, mental stress and energy intake and the risk of obesity in Genome Epidemiology Study. *Nutr Metab (Lond)*. 2016;13:38. doi:10.1186/s12986-016-0096-8
- 77.- Hager J, Clement K, Francke S, et al. A polymorphism in the 5' untranslated region of the human ob gene is associated with low leptin levels. *Int J Obes Relat Metab Disord*. 1998;22(3):200-205. doi:10.1038/sj.ijo.0800567
- 78.- Bienertova-Vasku J, Bienert P, Tomandl J, et al. No association of defined variability in leptin, leptin receptor, adiponectin, proopiomelanocortin and ghrelin gene with food preferences in the Czech population. *Nutr Neurosci*. 2008;11(1):2-8. doi:10.1179/147683008X301379
- 79.- Mammès O, Betoulle D, Aubert R, et al. Novel polymorphisms in the 5' region of the LEP gene: association with leptin levels and response to low-calorie diet in human obesity [published correction appears in *Diabetes* 2000;49(9):1617]. *Diabetes*. 1998;47(3):487-489. doi:10.2337/diabetes.47.3.487.
- 80.- Li WD, Reed DR, Lee JH, et al. Sequence variants in the 5' flanking region of the leptin gene are associated with obesity in women. *Ann Hum Genet*. 1999;63(Pt 3):227-234. doi:10.1046/j.1469-1809.1999.6330227.x

- 81.- Ben Ali S, Kallel A, Ftouhi B, et al. Association of G-2548A LEP polymorphism with plasma leptin levels in Tunisian obese patients. *Clin Biochem.* 2009;42(7-8):584-588. doi:10.1016/j.clinbiochem.2008.11.001
- 82.- Jiang Y, Wilk JB, Borecki I, et al. Common variants in the 5' region of the leptin gene are associated with body mass index in men from the National Heart, Lung, and Blood Institute Family Heart Study. *Am J Hum Genet.* 2004;75(2):220-230. doi:10.1086/422699
- 83.- Sun Q, Cornelis MC, Kraft P, et al. Genome-wide association study identifies polymorphisms in LEPR as determinants of plasma soluble leptin receptor levels [published correction appears in *Hum Mol Genet.* 2011;20(3):629]. *Hum Mol Genet.* 2010;19(9):1846-1855. doi:10.1093/hmg/ddq056
- 84.- Lu JF, Zhou Y, Huang GH, Jiang HX, Hu BL, Qin SY. Association of ADIPOQ polymorphisms with obesity risk: a meta-analysis. *Hum Immunol.* 2014;75(10):1062-1068. doi:10.1016/j.humimm.2014.09.004
- 85.- Goyenechea E, Collins LJ, Parra D, et al. The - 11391 G/A polymorphism of the adiponectin gene promoter is associated with metabolic syndrome traits and the outcome of an energy-restricted diet in obese subjects. *Horm Metab Res.* 2009;41(1):55-61. doi:10.1055/s-0028-1087204
- 86.- Li J, Jiang R, Cong X, Zhao Y. UCP2 gene polymorphisms in obesity and diabetes, and the role of UCP2 in cancer. *FEBS Lett.* 2019;593(18):2525-2534. doi:10.1002/1873-3468.13546.
- 87.- Jia JJ, Zhang X, Ge CR, Jois M. The polymorphisms of UCP2 and UCP3 genes associated with fat metabolism, obesity and diabetes. *Obes Rev.* 2009;10(5):519-526. doi:10.1111/j.1467-789X.2009.00569.x
- 88.- Ben Ali S, Ben Yahia F, Sediri Y, et al. Gender-specific effect of Pro12Ala polymorphism in peroxisome proliferator-activated receptor gamma-2 gene on obesity risk and leptin levels in a Tunisian population. *Clin Biochem.* 2009;42(16-17):1642-1647. doi:10.1016/j.clinbiochem.2009.08.019
- 89.- Rodrigues APS, Rosa LPS, Silveira EA. PPARG2 Pro12Ala polymorphism influences body composition changes in severely obese patients consuming extra virgin olive oil: a randomized clinical trial. *Nutr Metab (Lond).* 2018;15:52. doi:10.1186/s12986-018-0289-4

- 90.- Petr M, Stastny P, Zajac A, Tufano JJ, Maciejewska-Skrendo A. The Role of Peroxisome Proliferator-Activated Receptors and Their Transcriptional Coactivators Gene Variations in Human Trainability: A Systematic Review. *Int J Mol Sci.* 2018;19(5):1472. doi:10.3390/ijms19051472
- 91.- Leońska-Duniec A, Ahmetov II, Zmijewski P. Genetic variants influencing effectiveness of exercise training programmes in obesity - an overview of human studies. *Biol Sport.* 2016;33(3):207-214. doi:10.5604/20831862.1201052
- 92.- Corella D, Arnett DK, Tsai MY, et al. The -256T>C polymorphism in the apolipoprotein A-II gene promoter is associated with body mass index and food intake in the genetics of lipid lowering drugs and diet network study. *Clin Chem.* 2007;53(6):1144-1152. doi:10.1373/clinchem.2006.084863
- 93.- Pennacchio LA, Olivier M, Hubacek JA, et al. An apolipoprotein influencing triglycerides in humans and mice revealed by comparative sequencing. *Science.* 2001;294(5540):169-173. doi:10.1126/science.1064852
- 94.- Shu X, Chan J, Ryan RO, Forte TM. Apolipoprotein A-V association with intracellular lipid droplets. *J Lipid Res.* 2007;48(7):1445-1450. doi:10.1194/jlr.C700002-JLR200
- 95.- Herbert A, Gerry NP, McQueen MB, et al. A common genetic variant is associated with adult and childhood obesity. *Science.* 2006;312(5771):279-283. doi:10.1126/science.1124779
- 96.- Herbert A, Lenburg ME, Ulrich D, Gerry NP, Schlauch K, Christman MF. Open-access database of candidate associations from a genome-wide SNP scan of the Framingham Heart Study. *Nat Genet.* 2007;39(2):135-136. doi:10.1038/ng0207-135
- 97.- Frayling TM, Timpson NJ, Weedon MN, et al. A common variant in the FTO gene is associated with body mass index and predisposes to childhood and adult obesity. *Science.* 2007;316(5826):889-894. doi:10.1126/science.1141634
- 98.- Milagro FI, Campi3n J, Garc3a-D3az DF, Goyenechea E, Paternain L, Mart3nez JA. High fat diet-induced obesity modifies the methylation pattern of leptin promoter in rats. *J Physiol Biochem.* 2009;65(1):1-9. doi:10.1007/BF03165964.

- 99.- Funato H, Oda S, Yokofujita J, Igarashi H, Kuroda M. Fasting and high-fat diet alter histone deacetylase expression in the medial hypothalamus. *PLoS One*. 2011;6(4):e18950. Published 2011 Apr 15. doi:10.1371/journal.pone.0018950
- 100.- Drake AJ, Reynolds RM. Impact of maternal obesity on offspring obesity and cardiometabolic disease risk. *Reproduction*. 2010;140(3):387-398. doi:10.1530/REP-10-0077
- 101.- Plagemann A, Roepke K, Harder T, et al. Epigenetic malprogramming of the insulin receptor promoter due to developmental overfeeding. *J Perinat Med*. 2010;38(4):393-400. doi:10.1515/jpm.2010.051
- 102.- Vucetic Z, Kimmel J, Totoki K, Hollenbeck E, Reyes TM. Maternal high-fat diet alters methylation and gene expression of dopamine and opioid-related genes. *Endocrinology*. 2010;151(10):4756-4764. doi:10.1210/en.2010-0505
- 103.- Fillinger S, de la Garza L, Peltzer A, Kohlbacher O, Nahnsen S. Challenges of big data integration in the life sciences. *Anal Bioanal Chem*. 2019;411(26):6791-6800. doi:10.1007/s00216-019-02074-9
- 104.- Ngiam KY, Khor IW. Big data and machine learning algorithms for health-care delivery [published correction appears in *Lancet Oncol*. 2019;20(6):293]. *Lancet Oncol*. 2019;20(5):e262-e273. doi:10.1016/S1470-2045(19)30149-4.
- 105.- Wilkinson MD, Dumontier M, Aalbersberg IJ, et al. The FAIR Guiding Principles for scientific data management and stewardship [published correction appears in *Sci Data*. 2019 Mar 19;6(1):6]. *Sci Data*. 2016;3:160018. doi:10.1038/sdata.2016.18
- 106.- Ho DSW, Schierding W, Wake M, Saffery R, O'Sullivan J. Machine Learning SNP Based Prediction for Precision Medicine. *Front Genet*. 2019;10:267.. doi:10.3389/fgene.2019.00267.
- 107.- Visvikis D, Cheze Le Rest C, Jaouen V, Hatt M. Artificial intelligence, machine (deep) learning and radio(geno)mics: definitions and nuclear medicine imaging applications. *Eur J Nucl Med Mol Imaging*. 2019;46(13):2630-2637. doi:10.1007/s00259-019-04373-w
- 108.- Mitchell TM. The discipline of machine learning. *Mach. Learn*. 2006;17:1-7. <http://www.cs.cmu.edu/~tom/pubs/MachineLearning.pdf>

- 109.- Russell, S.N., P., *Artificial Intelligence: A Modern Approach*. 3rd edn. ed. 2013: Pearson, Harlow.
- 110.- Rodríguez-Pardo C, Segura A, Zamorano-León JJ, et al. Decision tree learning to predict overweight/obesity based on body mass index and gene polymorphisms. *Gene*. 2019;699:88-93. doi:10.1016/j.gene.2019.03.011
- 111.- Segura Fragoso A and Rius Mery G,. Cardiovascular risk factors in a rural population of Castilla-La Mancha. *Rev Esp Cardiol*, 1999;52(8):577-88.
- 112.- Modrego J, Azcona L, Martín-Palacios N, et al. Platelet content of nitric oxide synthase 3 phosphorylated at Serine 1177 is associated with the functional response of platelets to aspirin. *PLoS One*. 2013;8(12):e82574. doi:10.1371/journal.pone.0082574
- 113.- Zamorano-León JJ, Yañez R, Jaime G, et al. KCNH2 gene mutation: a potential link between epilepsy and long QT-2 syndrome. *J Neurogenet*. 2012;26(3-4):382-386. doi:10.3109/01677063.2012.674993
- 114.- Martínez-Hervas S, Mansego ML, de Marco G, et al. Polymorphisms of the UCP2 gene are associated with body fat distribution and risk of abdominal obesity in Spanish population. *Eur J Clin Invest*. 2012;42(2):171-178. doi:10.1111/j.1365-2362.2011.02570.x
- 115.- Argyropoulos G, Rankinen T, Neufeld DR, et al. A polymorphism in the human agouti-related protein is associated with late-onset obesity. *J Clin Endocrinol Metab*. 2002;87(9):4198-4202. doi:10.1210/jc.2002-011834
- 116.- Majithia AR, Flannick J, Shahinian P, et al. Rare variants in PPARG with decreased activity in adipocyte differentiation are associated with increased risk of type 2 diabetes [published correction appears in *Proc Natl Acad Sci U S A*. 2014;111(45):16225. Estrada, Karol [added]; Mercader, Josep [added]; MacArthur, Daniel [added]]. *Proc Natl Acad Sci U S A*. 2014;111(36):13127-13132. doi:10.1073/pnas.1410428111
- 117.- Price RA, Li WD, Zhao H. FTO gene SNPs associated with extreme obesity in cases, controls and extremely discordant sister pairs. *BMC Med Genet*. 2008;9:4. Published 2008 Jan 24. doi:10.1186/1471-2350-9-4
- 118.- Souren NY, Paulussen AD, Steyls A, et al. Common SNPs in LEP and LEPR associated with birth weight and type 2 diabetes-related metabolic risk factors in twins [published

correction appears in *Int J Obes (Lond)*. 2008 Aug;32(8):1335-6] [published correction appears in *Int J Obes (Lond)*. 2008 Nov;32(11):1745-6]. *Int J Obes (Lond)*. 2008;32(8):1233-1239. doi:10.1038/ijo.2008.68

119.- R Development Core Team, A Language and Environment for Statistical Computing. 2011.

120.- Barnholtz-Sloan JS, Guan X, Zeigler-Johnson C, Meropol NJ, Rebbeck TR. Decision tree-based modeling of androgen pathway genes and prostate cancer risk. *Cancer Epidemiol Biomarkers Prev*. 2011;20(6):1146-1155. doi:10.1158/1055-9965.EPI-10-0996

121.- Gauderman WJ. Sample size requirements for association studies of gene-gene interaction. *Am J Epidemiol*. 2002;155(5):478-484. doi:10.1093/aje/155.5.478

122.- Podgorelec V, Kokol P, Stiglic B, Rozman I. Decision trees: an overview and their use in medicine. *J Med Syst*. 2002;26(5):445-463. doi:10.1023/a:1016409317640

123.- Ollmann MM, Wilson BD, Yang YK, et al. Antagonism of central melanocortin receptors in vitro and in vivo by agouti-related protein [published correction appears in *Science* 1998;281(5383):1615]. *Science*. 1997;278(5335):135-138. doi:10.1126/science.278.5335.135

124.- Graham M, Shutter JR, Sarmiento U, Sarosi I, Stark KL. Overexpression of *Agrt* leads to obesity in transgenic mice. *Nat Genet*. 1997;17(3):273-274. doi:10.1038/ng1197-273

125.- Cowley MA, Pronchuk N, Fan W, Dinulescu DM, Colmers WF, Cone RD. Integration of NPY, AGRP, and melanocortin signals in the hypothalamic paraventricular nucleus: evidence of a cellular basis for the adipostat. *Neuron*. 1999;24(1):155-163. doi:10.1016/s0896-6273(00)80829-6

126.- Huszar D, Lynch CA, Fairchild-Huntress V, et al. Targeted disruption of the melanocortin-4 receptor results in obesity in mice. *Cell*. 1997;88(1):131-141. doi:10.1016/s0092-8674(00)81865-6

127.- Fan W, Boston BA, Kesterson RA, Hruby VJ, Cone RD. Role of melanocortinergic neurons in feeding and the agouti obesity syndrome. *Nature*. 1997;385(6612):165-168. doi:10.1038/385165a0

- 128.- Roselli-Rehfuss L, Mountjoy KG, Robbins LS, et al. Identification of a receptor for gamma melanotropin and other proopiomelanocortin peptides in the hypothalamus and limbic system. *Proc Natl Acad Sci U S A*. 1993;90(19):8856-8860. doi:10.1073/pnas.90.19.8856.
- 129.- Argyropoulos G, Rankinen T, Bai F, et al. The agouti-related protein and body fatness in humans. *Int J Obes Relat Metab Disord*. 2003;27(2):276-280. doi:10.1038/sj.ijo.802201
- 130.- Marks DL, Boucher N, Lanouette CM, et al. Ala67Thr polymorphism in the Agouti-related peptide gene is associated with inherited leanness in humans. *Am J Med Genet A*. 2004;126A(3):267-271. doi:10.1002/ajmg.a.20600
- 131.- Shutter JR, Graham M, Kinsey AC, Scully S, Lüthy R, Stark KL. Hypothalamic expression of ART, a novel gene related to agouti, is up-regulated in obese and diabetic mutant mice. *Genes Dev*. 1997;11(5):593-602. doi:10.1101/gad.11.5.593
- 132.- Vink T, Hinney A, van Elburg AA, et al. Association between an agouti-related protein gene polymorphism and anorexia nervosa. *Mol Psychiatry*. 2001;6(3):325-328. doi:10.1038/sj.mp.4000854
- 133.- Ilnytska O, Argyropoulos G. The role of the Agouti-Related Protein in energy balance regulation. *Cell Mol Life Sci*. 2008;65(17):2721-2731. doi:10.1007/s00018-008-8104-4
- 134.- Fleury C, Neverova M, Collins S, et al. Uncoupling protein-2: a novel gene linked to obesity and hyperinsulinemia. *Nat Genet*. 1997;15(3):269-272. doi:10.1038/ng0397-269
- 135.- Solanes G, Vidal-Puig A, Grujic D, Flier JS, Lowell BB. The human uncoupling protein-3 gene. Genomic structure, chromosomal localization, and genetic basis for short and long form transcripts. *J Biol Chem*. 1997;272(41):25433-25436. doi:10.1074/jbc.272.41.25433
- 136.- Chen HH, Lee WJ, Wang W, Huang MT, Lee YC, Pan WH. Ala55Val polymorphism on UCP2 gene predicts greater weight loss in morbidly obese patients undergoing gastric banding. *Obes Surg*. 2007;17(7):926-933. doi:10.1007/s11695-007-9171-6.

- 137.- Ochoa MC, Santos JL, Azcona C, et al. Association between obesity and insulin resistance with UCP2-UCP3 gene variants in Spanish children and adolescents. *Mol Genet Metab.* 2007;92(4):351-358. doi:10.1016/j.ymgme.2007.07.011
- 138.- Lee YH, Kim W, Yu BC, Park BL, Kim LH, Shin HD. Association of the ins/del polymorphisms of uncoupling protein 2 (UCP2) with BMI in a Korean population. *Biochem Biophys Res Commun.* 2008;371(4):767-771. doi:10.1016/j.bbrc.2008.04.144
- 139.- Zhang M, Wang M, Zhao ZT. Uncoupling protein 2 gene polymorphisms in association with overweight and obesity susceptibility: A meta-analysis. *Meta Gene.* 2014;2:143-159. doi:10.1016/j.mgene.2013.10.009
- 140.- Astrup A, Toubro S, Dalgaard LT, Urhammer SA, Sorensen TI, Pedersen O. Impact of the v/v 55 polymorphism of the uncoupling protein 2 gene on 24-h energy expenditure and substrate oxidation. *Int J Obes Relat Metab Disord.* 1999;23(10):1030-1034. doi:10.1038/sj.ijo.0801040.
- 141.- Walder K, Norman RA, Hanson RL, et al. Association between uncoupling protein polymorphisms (UCP2-UCP3) and energy metabolism/obesity in Pima indians. *Hum Mol Genet.* 1998;7(9):1431-1435. doi:10.1093/hmg/7.9.1431
- 142.- Klannemark M, Orho M, Groop L. No relationship between identified variants in the uncoupling protein 2 gene and energy expenditure. *Eur J Endocrinol.* 1998;139(2):217-223. doi:10.1530/eje.0.1390217
- 143.- Fritz T, Krämer DK, Karlsson HK, et al. Low-intensity exercise increases skeletal muscle protein expression of PPARdelta and UCP3 in type 2 diabetic patients. *Diabetes Metab Res Rev.* 2006;22(6):492-498. doi:10.1002/dmrr.656.
- 144.- Affourtit C, Brand MD. On the role of uncoupling protein-2 in pancreatic beta cells. *Biochim Biophys Acta.* 2008;1777(7-8):973-979. doi:10.1016/j.bbabi.2008.03.022
- 145.- Clapham JC, Arch JR, Chapman H, et al. Mice overexpressing human uncoupling protein-3 in skeletal muscle are hyperphagic and lean. *Nature.* 2000;406(6794):415-418. doi:10.1038/35019082
- 146.- Esterbauer H, Schneitler C, Oberkofler H, et al. A common polymorphism in the promoter of UCP2 is associated with decreased risk of obesity in middle-aged humans. *Nat Genet.* 2001;28(2):178-183. doi:10.1038/88911

- 147.- Ishigaki Y, Katagiri H, Yamada T, et al. Dissipating excess energy stored in the liver is a potential treatment strategy for diabetes associated with obesity. *Diabetes*. 2005;54(2):322-332. doi:10.2337/diabetes.54.2.322
- 148.- Yu X, Jacobs DR Jr, Schreiner PJ, Gross MD, Steffes MW, Fornage M. The uncoupling protein 2 Ala55Val polymorphism is associated with diabetes mellitus: the CARDIA study. *Clin Chem*. 2005;51(8):1451-1456. doi:10.1373/clinchem.2004.044859
- 149.- de Souza BM, Brondani LA, Bouças AP, et al. Associations between UCP1 -3826A/G, UCP2 -866G/A, Ala55Val and Ins/Del, and UCP3 -55C/T polymorphisms and susceptibility to type 2 diabetes mellitus: case-control study and meta-analysis. *PLoS One*. 2013;8(1):e54259. doi:10.1371/journal.pone.0054259
- 150.- Xiu LL, Weng JP, Sui Y, Wang J, Yan JH, Huang ZM. *Zhonghua Yi Xue Za Zhi*. 2004;84(5):375-379.
- 151.- Astrup A, Toubro S, Dalgaard LT, Urhammer SA, Sorensen TI, Pedersen O. Impact of the v/v 55 polymorphism of the uncoupling protein 2 gene on 24-h energy expenditure and substrate oxidation. *Int J Obes Relat Metab Disord*. 1999;23(10):1030-1034. doi:10.1038/sj.ijo.0801040
- 152.- Rosmond R, Bouchard C, Björntorp P. Lack of association between the uncoupling protein-2 Ala55Val gene polymorphism and phenotypic features of the Metabolic Syndrome. *Biochim Biophys Acta*. 2002;1588(2):103-105. doi:10.1016/s0925-4439(02)00152-7
- 153.- Tarnopolsky M. Gender differences in metabolism: practical and nutritional implications. Ed. Boca Raton: CRC Press; 1999.
- 154.- Day N, Oakes S, Luben R, et al. EPIC-Norfolk: study design and characteristics of the cohort. *European Prospective Investigation of Cancer*. *Br J Cancer*. 1999;80 Suppl 1:95-103.
- 155.- Wareham NJ, Byrne CD, Williams R, Day NE, Hales CN. Fasting proinsulin concentrations predict the development of type 2 diabetes. *Diabetes Care*. 1999;22(2):262-270. doi:10.2337/diacare.22.2.262
- 156.- Orkunoglu-Suer FE, Gordish-Dressman H, Clarkson PM, et al. INSIG2 gene polymorphism is associated with increased subcutaneous fat in women and poor

response to resistance training in men. *BMC Med Genet.* 2008;9:117. Published 2008 Dec 23. doi:10.1186/1471-2350-9-117

157.- Tabara Y, Kawamoto R, Osawa H, et al. No association between INSIG2 Gene rs7566605 polymorphism and being overweight in Japanese population. *Obesity (Silver Spring).* 2008;16(1):211-215. doi:10.1038/oby.2007.25

158.- Dina C, Meyre D, Samson C, et al. Comment on "A common genetic variant is associated with adult and childhood obesity". *Science.* 2007;315(5809):187. doi:10.1126/science.1129402

159.- Roskopf D, Bornhorst A, Rimbach C, et al. Comment on "A common genetic variant is associated with adult and childhood obesity". *Science.* 2007;315(5809):187. doi:10.1126/science.1130571

160.- Reinehr T, Hinney A, Nguyen TT, Hebebrand J. Evidence of an influence of a polymorphism near the INSIG2 on weight loss during a lifestyle intervention in obese children and adolescents. *Diabetes.* 2008;57(3):623-626. doi:10.2337/db07-0408

9. ANEXO I

Publicación generada con los resultados del trabajo de la tesis doctoral en revista “Gene”, indexada en JCR (Impact factor: 3.331, Ranking 61/174. Q2).



Research paper

Decision tree learning to predict overweight/obesity based on body mass index and gene polymorphisms



Carlos Rodríguez-Pardo^a, Antonio Segura^b, José J. Zamorano-León^c, Cristina Martínez-Santos^c, David Martínez^c, Luis Collado-Yurrita^d, Manel Giner^e, José M. García-García^f, José M. Rodríguez-Pardo^g, Antonio López-Farre^{d,*}

^a Applied Artificial Intelligence Team, Universidad Carlos III de Madrid, Madrid, Spain

^b Health Science Institute, Talavera de la Reina, Toledo, Spain

^c Department of Public Health and Maternal and Child Health, School of Medicine, Universidad Complutense, Madrid, Spain

^d Department of Medicine, School of Medicine, Universidad Complutense, Madrid, Spain

^e Department of Surgery, School of Medicine, Universidad Complutense, Madrid, Spain

^f Physical Activity and Sport Sciences Department, School of Sport Sciences, Universidad Castilla-La Mancha, Toledo, Spain

^g Life and Health Unit, Universidad Carlos III, Madrid, Spain

ARTICLE INFO

Keywords:

Decision tree
Overweight
Obesity
Genotype
Single nucleotide polymorphism
Body mass index

ABSTRACT

The new technologies for data analysis, such as decision tree learning, may help to predict the risk of developing diseases. The aim of the present work was to develop a pilot decision tree learning to predict overweight/obesity based on the combination of six single nucleotide polymorphisms (SNP) located in feeding-associated genes. Genotype study was performed in 151 healthy individuals, who were anonymized and randomly selected from the TALAVERA study. The decision tree analysis was performed using the R package rpart. The learning process was stopped when 15 or less observation was found in a node. The participant group consisted of 78 men and 73 women, who 100 individuals showed body mass index (BMI) ≥ 25 kg/m² and 51 BMI < 25 kg/m². Chi-square analysis revealed that individuals with BMI ≥ 25 kg/m² showed higher frequency of the allelic variation Ala67Ala in AgRP rs5030980 with respect to those with BMI < 25 kg/m². However, the variant Thr67Ala in AgRP rs5030980 was the most frequently found in individuals with BMI < 25 kg/m². There were no statistical differences in the other analyzed SNPs. Decision tree learning revealed that carriers of the allelic variants AgRP (rs5030980) Ala67Ala, ADRB2 (rs1042714) Gln27Glu or Glu27Glu, INSIG2 (rs7566605) 73 + 9802 with CC or GG genotypes and PPARG (rs1801282) with the allelic variants of Ala12Ala or Pro12Pro, will most likely develop overweight/obesity (BMI ≥ 25 kg/m²). Moreover, the decision tree learning indicated that age and gender may change the developed three decision learning associated with overweight/obesity development. The present work should be considered as a pilot demonstrative study to reinforce the broad field of application of new data analysis technologies, such as decision tree learning, as useful tools for diseases prediction. This technology may achieve a potential applicability in the design of early strategies to prevent overweight/obesity.

1. Introduction

Prevalence of overweight/obesity has reached epidemic levels in Western societies. However, overweight and obesity are considered major modifiable health risk factors that contribute to a global burden of chronic diseases. Indeed, high body mass index (BMI) has been associated with increased risk of morbidity and mortality diseases risk

(Adams et al., 2006; Must et al., 1999). Moreover, overweight and obesity also exert direct and indirect economic impacts, increasing hospital care, pharmaceuticals and productivity costs (Tremmel et al., 2017; Bhattacharya and Bundorf, 2009). Furthermore, overweight/obesity also affects physical and mental health conditions which equally have an impact on reducing productivity (Nigatu et al., 2016).

Several factors have been involved in overweight and obesity.

Abbreviations: SNP, single nucleotide polymorphism; BMI, body mass index; GWA, genomic wide association; INSIG2, insulin-induced gene 2; PPARG, peroxisome proliferator activated receptor gamma; ADRB2, beta-2 adrenergic receptor; ADRB3, beta-3 adrenergic receptor; AGRP, agouti-related protein; UCP2, Mitochondrial uncoupling protein 2

* Corresponding author at: Medicine Department, School of Medicine, Universidad Complutense, Plaza Ramón y Cajal. SN, Madrid 28040, Spain.

E-mail address: ajlf@telefonica.net (A. López-Farre).

<https://doi.org/10.1016/j.gene.2019.03.011>

Received 11 July 2018; Received in revised form 6 March 2019; Accepted 7 March 2019

Available online 08 March 2019

0378-1119/ © 2019 Elsevier B.V. All rights reserved.

However, one of the main attributed causes is an imbalance between the energy intake and the energy expenditure (Romieu et al., 2017; Davison and Birch, 2001). Therefore, nutritional approaches have been traditionally focused on the assumption that all individuals have the same nutritional requirements. However, inter-individual variability in response to a diet is clear, where genetic predisposition may have a relevant role. In this regard, there is emerging evidence showing the role of specific variants, single nucleotide polymorphisms (SNP), in genes involved in the efficiency of the metabolism of specific macronutrients (Garaulet et al., 2011; Kaulfers et al., 2015; Nicoletti et al., 2017). Indeed, large-scale genome-wide association studies (GWAS) have identified a number of selected SNP with variations in the BMI (Yasukochi et al., 2018; Akiyama et al., 2017; Song et al., 2017). Nevertheless, predictable capacity of genome scans is widely questioned by several researchers, since the basic approach of GWAS is to evaluate the association between each genotyped marker and a phenotype of interest in a large number of individuals: In this sense, GWAS technique does not have the ability to identify possible direct relationships among SNPs from different genes to predict a phenotype. Indeed, the GWAS strategy explains < 2% of the inter-individual BMI variation, and < 1% of the inter-individual waist-to-hip ratio variation (Lindgren et al., 2009; Willer et al., 2009). Interestingly, genetic epistasis, defined as a masking relationship between two loci (statistical interaction occurs when the effect of one locus depends on the genotypes of another loci), may have a relevant role to predict individual genetic probability to develop overweight/obesity (De et al., 2015). However, in our knowledge, the possible relationships between different SNPs to predict the probability of developing overweight/obesity prediction have not been sufficiently assessed.

Machine learning is an advanced computer science field that enables machines to learn patterns from data. In this regard, decision trees are decision structures that recursively subdivide the data to classify each sample depending on the values of the variables with which the samples are defined. (Russell and Norvig, 2013; Mitchell, 2006). In the field of genetics, decision trees learning has started to be used to annotate a wide variety of genomic sequence elements such as splice sites, promoters, enhancers, positioned nucleosomes, etc. (Degroevie et al., 2002; Bucher, 1990; Heintzman et al., 2007; Segal et al., 2006). However, in our knowledge decision tree learning has not been used to predict overweight/obesity using specific SNPs associated with nutrition. Therefore, we here developed a pilot decision tree analysis to predict normal and overweight/obese classes based on the combination of SNPs located in feeding-associated genes.

2. Material and methods

2.1. Study population

As we previously reported in greater detail (Segura Fragoso and Rius Mery, 1999), between 1994 and 1996, individuals older than 18 years living in the area of the city of Talavera de la Reina, Toledo, Spain, were invited to participate in the TALAVERA study. It is a rural population (approximately 90.000 inhabitants) distributed in municipalities of < 5.000 inhabitants each. At that time, 1.500 individuals were recruited. The present study was performed in 151 anonymized and randomly selected individuals from the TALAVERA study. Participants with any type of chronic disease and hospitalized by any cause at any time before inclusion were excluded. BMI was measured as weight in kg per height in squared meters. Participants were classified into two groups according to BMI values from World Health Organization (WHO). Group one was constituted by participants with BMI < 25 group ($n = 51$) that included participants with BMI values in the BMI range 18.50–24.99 kg/m². Group two was constituted by participants with BMI ≥ 25 group ($n = 100$) including individuals with either BMI values between 25.00 and 29.99 kg/m² ($n = 55$) and individuals with BMI values higher than 29.99 kg/m² ($n = 45$). The research was

Table 1
Hardy–Weinberg equilibrium analysis of the six single nucleotide polymorphisms.

Genetic alteration	HWE <i>P</i> value	Minor allelic frequency	Minor allele	Ancestral allele
UCP2 (rs660339)	0.665	0.421	A	G
AgRP (rs5030980)	< 0.001	0.252	T	C
PPARG (rs1801282)	0.905	0.099	G	C
ADRB3 (rs4994)	0.814	0.050	G	G
ADRB2 (rs1042714)	0.649	0.351	G	G
INSIG2 (rs7566605)	0.132	0.371	C	G

approved by the Ethics Local Committee of the Hospital Clínico San Carlos (protocol # C.I. 14/442-E).

2.2. SNP selection and genotyping

Due to limited resources availability, we selected six SNPs previously reported as potentially associated with BMI and/or macronutrients metabolism (Martinez-Hervas et al., 2012; Argyropoulos et al., 2002; Ben Ali et al., 2009; Takeuchi et al., 2012; Ruiz et al., 2011; Herbert et al., 2006; Price et al., 2008; Souren et al., 2008) (Table 1).

Blood samples were collected for DNA extraction. As we previously reported (Modrego et al., 2013; Zamorano-León et al., 2012), genomic DNA was extracted from the peripheral blood according to standard phenol–chloroform methods. We genotyped SNPs using the TaqMan assay. The sample DNA was amplified by PCR on a GeneAmp PCR System 9700 thermal cycler (Applied Biosystems) following the manufacturer's recommendations. Genotypes were differentiated by analyzing the fluorescence levels of PCR products using an ABI PRISM 7900HT Sequence Detector (Applied Biosystems). Genotyping was blindly performed and all sequences were carried out in the central services of Madrid Science Park. Table 1 shows the analyzed SNPs.

2.3. Machine learning and decision trees

To perform the decision trees, the R package rpart was used (<https://CRAN.R-project.org/package=rpart>), and learning was stopped when 15 or less observations were found in a node, to avoid the construction of too complex trees (Development Core, 2011). The purpose of using decision trees in this task is mostly exploratory, which means that there was no division between training and test data, and, consequently, common accuracy metrics like AUC or binary loss will not be reported.

2.4. Statistical analysis

Age was compared using the t-Student test. The allele frequencies were estimated by direct gene counting and were analyzed for Hardy–Weinberg equilibrium (HWE) using a Chi-square good-of-fit test. Distribution of genotypes in two groups was analyzed using Chi-square test. Data are given as means \pm EEM, unless otherwise indicated. $P < 0.05$ was considered significant. Statistical analysis was performed using the SPSS program (version 22.0).

3. Results

3.1. Genotyping results

The average age of the participants included was 55.1 ± 1.34 years old, where 78 were men (mean age 55.64 ± 1.82) and 73 women (mean age 54.64 ± 2.01). As Table 2 shows, age was slightly higher in the individuals with BMI ≥ 25 kg/m² with respect to those with BMI < 25 kg/m² but reached statistical significance (Table 2).

The genotype frequencies for the tested polymorphisms and genes in

Table 2
Genotype frequencies of the analyzed single nucleotide polymorphisms according to body mass index.

Genes		BMI < 25 group (N = 51)	BMI ≥ 25 group (N = 100)	P value
Age (years)		50.75 ± 2.70	57.32 ± 1.45	0.03
Genes		Genotypes frequencies (cases/%)		
UCP2 (rs660339)	AA (Val55Val)	10/19.6%	14/14.0%	0.66
	AG (Val55Ala)	26/51.0%	53/53.0%	
	GG (Ala55Ala)	15/29.4%	33/33.0%	
AgRP (rs5030980)	CC (Ala67Ala)	18/35.3%	57/57.0%	0.01
	CT (Ala67Thr)	33/64.7%	43/43.0%	
	TT (Thr67Thr)	0/0.0%	0/0.0%	
PPARG (rs1801282)	CC (Pro12Pro)	39/76.5%	83/83.0%	0.28
	CG (Pro12Ala)	11/21.6%	17/17.0%	
	GG (Ala12Ala)	1/2.0%	0/0.0%	
ADRB3 (rs4994)	AA (Trp64Trp)	45/88.2%	91/91.0%	0.59
	AG (Trp64Arg)	6/11.8%	9/9.0%	
	GG (Arg64Arg)	0/0.0%	0/0.0%	
ADRB2 (rs1042714)	CC (Gln67Gln)	23/45.1%	38/38.0%	0.05
	CG (Gln67Glu)	19/37.3%	55/55.0%	
	GG (Glu67Glu)	9/17.6%	7/7.0%	
INSIG2 (rs7566605)	CC (no coding region)	4/7.8%	11/11.0%	0.73
	CG (no coding region)	27/52.9%	55/55.0%	
	GG (no coding region)	20/39.2%	34/34.0%	

the study population are shown in Table 1. The genotype frequencies of five gene variants were consistent with the Hardy-Weinberg equilibrium (*p*-values ranging from 0.132 to 0.905). Only agouti-related protein (AgRP) variant deviated significantly from Hardy-Weinberg equilibrium, it could be attributed to type I error.

As Table 2 shows, Chi-square analysis demonstrated statistical differences between individuals in the allelic distribution of the AgRP gene polymorphism rs5030980. Indeed, the data suggested higher percentage of individuals with IBM ≥ 25 kg/m² showing the allele variant Ala67Ala with respect to the individuals with BMI < 25 kg/m². However, the allelic variant Thr67Ala in AgRP (rs5030980) tended to be more frequent in individuals with BMI < 25 kg/m² than those with BMI > 25 kg/m². The statistical analysis also showed that the allelic distribution of the beta-2 adrenergic receptor (ADRB2) gene polymorphism rs1042714 between the individuals showing BMI < 25 kg/m² and those with BMI ≥ 25 kg/m² tended to be difficult (Table 2). Indeed, the presence of the Gln27Gln variant in the population with BMI < 25 kg/m² was more frequent as compared to individuals with BMI ≥ 25 kg/m² (Table 2). There were no statistical differences in the other UCP polymorphisms (Table 2).

3.2. Decision tree learning associating BMI with gene polymorphisms

As mentioned above, 34% of the study population showed BMI < 25 kg/m² and 66% have BMI ≥ 25 kg/m² (Fig. 1). The learning system divided the 151-included individuals into two branches regarding to the Ala67Thr polymorphism (rs5030980) in the AgRP gene (Fig. 1). 50.7% of the total included population showed the Ala67Thr polymorphism (CT alleles). 56.6% of them had BMI ≥ 25 kg/m² and 43.4% showed BMI < 25 kg/m² (Fig. 1). The remaining 49.3% of the total population showed non Ala67Thr variant in the AgRP rs5030980 polymorphism and 75.7% of these individuals were associated with BMI ≥ 25 kg/m² (Fig. 1).

A new descendent branch appeared from the descendent node of the AgRP Ala67Thr polymorphism rs5030980 (CT allelic variant) (Fig. 1). This new branch indicated that in the population showing the variant Ala67Thr in AgRP rs5030980 (50.7% of total population), only 6.7% of them showed the variant Gln27Gln (allele GG) in ADRB2 rs1042714.

However, 70% of these individuals had normal weight, BMI < 25 kg/m² (Fig. 1). Moreover, 44% of the population showing the variant Ala67Thr in AgRP rs5030980, also showed the allelic variants GC or CC in ADRB2 rs1042714 (Fig. 1).

A new branch derived from the node of this later population showing the allelic variants GC or CC in the ADRB2 polymorphism rs1042714, indicated that if the insulin-induced gene 2 (INSIG2) polymorphism rs7566605 had the allelic variant GC (52.9% of individuals with ADRB2 genetic variants GC or CC), individuals underwent BMI ≥ 25 kg/m² (Fig. 1). Furthermore, if the UCP2 polymorphism (rs660309) was GG, 71.4% of these later individual will develop BMI ≥ 25 kg/m² (Fig. 1).

Going back to the INSIG2 node, within 44% of the population showing the allelic variant GC or CC in the ADRB2 polymorphism rs1042714 (Fig. 1), 21.3% of them also showed the allelic variants CC or GG in the INSIG2 gene polymorphism rs7566605. In this population, 68.8% individuals have BMI ≥ 25 kg/m². Descending along this branch, another new node was identified in the decision tree learning, suggesting that if the peroxisome proliferator activated receptor gamma encoding gene (PPARG) polymorphism rs1801282 has the allelic variant GC (4.7% of the population with INSIG2 other than the CG allelic variant in rs7566605), 42.9% of these individuals will show overweight/obesity (BMI ≥ 25 kg/m²). However, if this PPARG polymorphism showed the allelic variants CC or GG, 76.0% of these individuals will show overweight/obesity (Fig. 1).

3.3. Impact of age and gender on the decision tree learning

This decision tree learning can be complicated, and potentially increases its accuracy by adding other factors, such as age and gender. Including these two factors, the learning system divided the population into two branches according to individuals ≥ 38.5 years or younger (Fig. 2).

As Fig. 2 shows, only 16% of the study population were under 38.5 years old. As expected, 75% of these younger individuals showed BMI < 25 kg/m². On the other hand, 84% of the total population were over 38.5 years old and, 73.8% of them, showed BMI ≥ 25 kg/m² (Fig. 2).

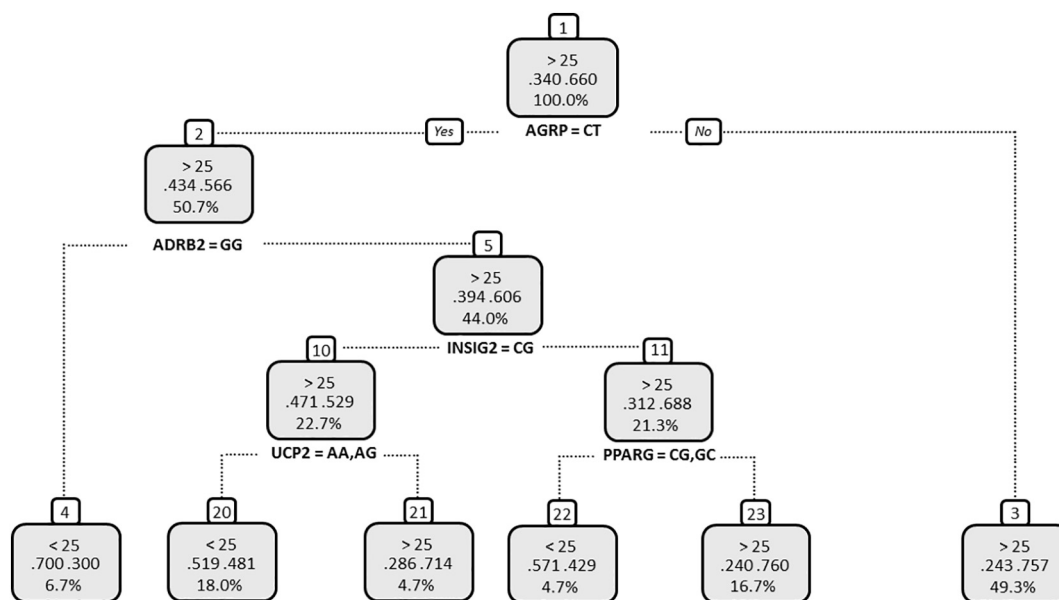


Fig. 1. Decision tree model using Body Mass Index (BMI) and single nucleotide polymorphisms (SNPs) Related to genes associated with feeding and obesity.

From the descendent node of ≥ 38.5 years, genotyping differences promoted a new node. In this node, 39,3% of the total population with age over 38.5 years old showed the allelic variant CT in AgRP gene polymorphism rs5030980, and 64.4% of them have BMI ≥ 25 kg/m² (Fig. 2). In addition, 82% individuals with age over 38.5 years demonstrated BMI ≥ 25 kg/m² when AgRP allelic variants in the rs5030980 polymorphism were different to CT (Ala67Thr) (Fig. 2).

Considering age between 38.5 and ≥ 70.5 years old, the decision tree learning added a new branch showing the AGRP gene polymorphism rs5030980 allelic variant CT (Fig. 2). The decision tree learning determines that 10% of individuals showing the allelic variant CT in the rs5030980 AgRP polymorphism are between 38.5 and ≥ 70.5 years old and 60% of them had BMI < 25 kg/m² (Fig. 2). Within individuals showing BMI < 25 kg/m², 4% of them showed the allelic variant AA in the UCP2 gene polymorphism rs660339, with 83.3% from this 4%, having BMI < 25 kg/m² (Fig. 2).

On the other hand, 29.3% of individuals showing the allelic variant CT in the AgRP polymorphism rs5030980 were not in the age interval of 38.5- ≥ 70.5 years old and 72.7% of these individuals showed BMI ≥ 25 kg/m² (Fig. 2). Furthermore, 17.3% of these individuals were men and 84.6% of them had BMI ≥ 25 kg/m² (Fig. 2). 12% of the individuals outside the interval 38.5- ≥ 70.5 years old and showing the allelic variant CT in AgRP rs5030980 were women. 55.6% of these women showed BMI ≥ 25 Kg/m². Moreover, 7.3% of the women with age other than that included between 38.8- ≥ 70.5 years old and showing the allelic variant CT in rs5030980 AgRP and the allelic variant CC or CG in INSIG2 rs7566605. Within this female population, 63.5% of them showed BMI < 25 kg/m²(Fig. 2). 4.7% were carriers of the allelic variant GG in INSIG2 gene polymorphism rs7566605. From them, 84.6% have BMI ≥ 25 kg/m² (Fig. 2).

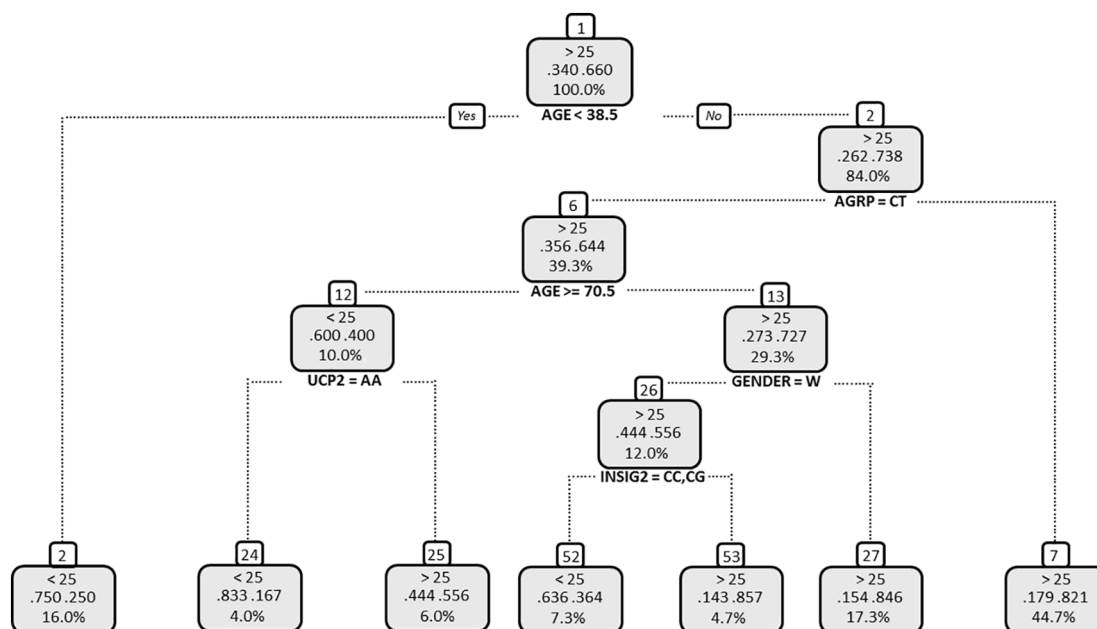


Fig. 2. Impact of age and gender in the decision tree model.

4. Discussion

This is a pioneer pilot study that proposes decision trees to predict BMI using SNPs associated with both feeding and BMI-related genes.

The decision tree learning identifies relationships between variables and associations with risks that may not be identified by traditional epidemiologic analysis techniques (Barnholtz-Sloan et al., 2011). The present work may serve as an example, as the information obtained from the more conventional statistical genetic analysis in the study was less than that derived from the decision trees learning. In this sense, a relevant strength of the decision tree in the analysis of genetic information is to establish relationships among different polymorphisms as well as between these genetic variations and other factors such as age and gender. Consequently, although in biomedical research some works have started to use decision tree learning with biomedical biomarkers (Podgorelec et al., 2002), in our knowledge, this is one of the first works using decision tree learning associating genetic information, related to feeding-associated genes, with overweight/obesity prediction.

The first node that appeared in the decision tree learning suggested that in our study population, the absence of Ala67Thr AgRP polymorphism is more frequently associated with BMI ≥ 25 kg/m². AgRP gene encodes an antagonist of the melanocortin-3 and -4 and regulates hypothalamic control of feeding behaviours (Ollmann et al., 1997). There are data associating AgRP-encoding gene with obesity (Argyropoulos et al., 2002; Argyropoulos et al., 2003; Marks et al., 2004). Indeed, hypothalamic AgRP is elevated in obese and diabetic mice while transgenic mice overexpressing AgRP are obese (Shutter et al., 1997). However, in humans, the Ala67Thr variant in rs67483042 AgRP has been associated with lower body weight and even with anorexia nervosa (Marks et al., 2004; Vink et al., 2001).

Descending in the decision tree, it is possible to draw the involvement of more genetic polymorphisms, the combination of which will allow us to define with higher accuracy the genetic characteristics of an individual with and without overweight/obesity. Accordingly, a possible combination arising from the first decision tree learning was a combination of genetic polymorphisms with the following allelic variants: AgRP (rs5030980) Ala67Ala, ADRB2 (rs1042714) Gln27Glu or Glu27Glu, INSIG2 (rs7566605) 73 + 9802 the CC or GG genotypes and PPARG (rs1801282) Ala12Ala or Pro12Pro probably identifies individuals with BMI ≥ 25 kg/m².

The decision tree learning could gain accuracy adding new factors. In the present study, the decision tree learning was then performed adding age and gender factors since both of them are recognized to influence overweight/obesity development (Song et al., 2010). In this regard, age was found significantly higher in the individuals with BMI ≥ 25 kg/m² in the analyzed population.

In the study population, the first branch raised from the inclusion of two new factors, age and gender factors, revealed that individuals over 38.5 years old will more likely be overweight/obese. Within the individuals over 38.5 years old the allelic variant CT in rs5030980 AgRP gene defined a new descending branch. Indeed, in the population over 38.5 years old, absence of the CT allelic variant (Ala67Thr) increased the risk for BMI ≥ 25 kg/m². Interestingly, a previous study has suggested that Ala67Thr polymorphism was consistently associated with BMI reduction, in the case of the parents, but not so in the case of the offspring (Ilnytska and Argyropoulos, 2008). It may suggest, that in some way, the Ala67Thr genotype could be exerting its effects in an age-dependent fashion.

The influence of age on the decision tree learning was even narrower in a new descending branch. This node predicted overweight/obesity in individuals between 38.5 and 70.5 years old and the allelic variant CT in the AgRP polymorphism rs5030980. Within them, the decision tree learning identified another gen allelic variant, AA in UCP2 gene polymorphism rs660339. This node points out that 83.3% of individuals will have BMI < 25 kg/m² if the following conditions occur: to be aged between 38.5 and 70.5 years old, the allelic variant CT in

AgRP gene polymorphism rs5030980 and the allelic variant AA in UCP2 gene polymorphism rs660339. Therefore, the presence of the Ala allelic in rs660339 UCP-2 polymorphism variant may increase protection against overweight/obesity in our population. Accordingly, in a meta-analysis, the allelic variant Val in this UCP2 polymorphism Ala55Val was associated with an increased risk of overweight/obesity (Zhang et al., 2014). Moreover, the allelic variant 55Val/Val in the UCP2 polymorphism Ala55Val was associated with higher incidence of type-2 diabetes mellitus in Caucasians and African-Americans (Yu et al., 2005). This observation was also reported by other authors (de Souza et al., 2013). Indeed, individuals with the Val/Val genotype in the UCP2 Ala55Val polymorphism appear to have a lower energy expenditure and increased susceptibility to obesity and type 2 diabetes mellitus (Xiu et al., 2004; Astrup et al., 1999). Other studies have reported that this polymorphism is not individually associated with basal metabolic rate, metabolic syndrome, BMI, obesity, insulin secretion or type 2 diabetes mellitus (Rosmond et al., 2002; Lee et al., 2008). The inconsistencies about this polymorphism reported in the results may be partially explained by different factors including sample size, ethnicity, age and sex. In this sense, the decision tree learning shows the advantage of interacting age and clustering with other gene polymorphisms.

Adult body size and body shape differ substantially between men and women and change over time. For this reason, it is important to establish and identify genetic polymorphisms that may be differently associated with overweight/obesity, regarding sex differences. The decision tree learning algorithm identified a percentage of women that they will have BMI ≥ 25 kg/m² when they carry the allelic clustering variants CT in the AgRP polymorphism rs5030980 and GG in INSIG2 gene polymorphism rs7566605 and age was not included in the interval 38.5–70.5 years old. Interestingly, a study suggested that women with a copy of the C allele had higher levels of baseline subcutaneous fat, but men did not show such association (Orkunoglu-Suer et al., 2008). However, another study in Japanese population suggested a lack of association between SNP rs7566605 in INSIG2-encoded gene and being overweight (in the middle-aged and elderly population) (Tabara et al., 2008). As mentioned above, factors such as ethnicity, age and clustering with other SNPs probably influence deeply the impact of specific SNPs on overweight/obesity development.

As conclusion, although several limitations can be obviously established in this pilot study, including the limited sample size, which also determined that the group showing BMI ≥ 25 kg/m² could not be divided into those individuals showing overweight from those with obesity. Therefore, this work should be considered as a pilot demonstrative study to support the broad field of the application of new data analysis techniques like decision tree learning as useful tools for diseases prediction. In this sense, the possibility of predicting if an individual will have greater susceptibility to suffer overweight/obesity according to their genotype, age and gender factors, among other factors, may be of great applicability in designing early strategies to prevent overweight/obesity.

Acknowledgments

Authors thank Begoña Larrea for her editorial assistance. Jose J. Zamorano-León, José M. García-García and Antonio López-Farré are members of the interuniversity collaborative research group AKANTHOS.

Funding source

This work was supported by a grant from the Comunidad de Madrid (GenObIA-CM, B2017/BMD-3773) and co-supported by Structural Funds of the European Union.

Competing interests

The authors declare that there are no competing financial interests.

References

- Adams, K.F., Schatzkin, A., Harris, T.B., Kipnis, V., Mouw, T., Ballard-Barbash, R., et al., 2006. Overweight, obesity, and mortality in a large prospective cohort of persons 50 to 71 years old. *New Engl J Med* 355, 763–778.
- Akiyama, M., Okada, Y., Kanai, M., Takahashi, A., Momozawa, Y., Ikeda, M., et al., 2017. Genome-wide association study identifies 112 new loci for body mass index in the Japanese population. *Nat. Genet.* 49, 1458–1467.
- Argyropoulos, G., Rankinen, T., Neufeld, D.R., Rice, T., Province, M.A., Leon, A.S., et al., 2002. A polymorphism in the human agouti-related protein is associated with late-onset obesity. *J. Clin. Endocrinol. Metab.* 87, 4198–4202.
- Argyropoulos, G., Rankinen, T., Bai, F., Rice, T., Province, M.A., Leon, A.S., et al., 2003. The agouti-related protein and body fatness in humans. *Int. J. Obes. Relat. Metab. Disord.* 27, 276–280.
- Astrup, A., Toubro, S., Dalgaard, L.T., Urhammer, S.A., Sorensen, T.I., Pedersen, O., 1999. Impact of the v/v 55 polymorphism of the uncoupling protein 2 gene on 24-h energy expenditure and substrate oxidation. *Int. J. Obes. Relat. Metab. Disord.* 23, 1030–1034.
- Barnholtz-Sloan, J.S., Guan, X., Zeigler-Johnson, C., Meropol, N.J., Rebbeck, T.R., 2011. Decision tree-based modeling of androgen pathway genes and prostate cancer risk. *Cancer Epidemiol. Biomark. Prev.* 20, 1146–1155.
- Ben Ali, S., Ben Yahia, F., Sediri, Y., Kallel, A., Ftouhi, B., Feki, M., et al., 2009. Gender-specific effect of Pro12Ala polymorphism in peroxisome proliferator-activated receptor gamma-2 gene on obesity risk and leptin levels in a Tunisian population. *Clin. Biochem.* 42, 1642–1647.
- Bhattacharya, J., Bundorf, M.K., 2009. The incidence of the healthcare costs of obesity. *J. Health Econ.* 28, 649–658.
- Bucher, P., 1990. Weight matrix description of four eukaryotic RNA polymerase II promoter elements derived from 502 unrelated promoter sequences. *J. Mol. Biol.* 4, 563–578.
- Davison, K.K., Birch, L.L., 2001. Childhood overweight: a contextual model and recommendations for future research. *Obes. Rev.* 2, 159–171.
- de Souza, B.M., Brondani, L.A., Bouças, A.P., Sortica, D.A., Kramer, C.K., Canani, L.H., et al., 2013. Associations between UCP1-3826A/G, UCP2-866G/A, Ala55Val and Ins/Del, and UCP3-55C/T polymorphisms and susceptibility to type 2 diabetes mellitus: case-control study and meta-analysis. *PLoS One* 8, e54259.
- De, R., Hu, T., Moore, J.H., Gilbert-Diamond, D., 2015. Characterizing gene-gene interactions in a statistical epistasis network of twelve candidate genes for obesity. *BioData Mining* 8, 45.
- Degroeve, S., Baets, B.D., de Peer, Y.V., Rouz, P., 2002. Feature subset selection for splice site prediction. *Bioinformatics* 18, 75–83.
- Garaulet, M., Smith, C.E., Hernández-González, T., Lee, Y.-C., Ordovás, J.M., 2011. PPARγ Pro12Ala interacts with fat intake for obesity and weight loss in a behavioural treatment based on the Mediterranean diet. *Mol. Nutr. Food Res.* 55, 1771–1779.
- Heintzman, N.D., Stuart, R.K., Hon, G., Fu, Y., Ching, C.W., Hawkins, R.D., et al., 2007. Distinct and predictive chromatin signatures of transcriptional promoters and enhancers in the human genome. *Nat. Genet.* 39, 311–318.
- Herbert, A., Gerry, N.P., McQueen, M.B., Heid, I.M., Pfeufer, A., Illig, T., et al., 2006. A common genetic variant is associated with adult and childhood obesity. *Science* 312, 279–283.
- Ilnytska, O., Argyropoulos, I., 2008. The role of the agouti-related protein in energy balance regulation. *Cell. Mol. Life Sci.* 65, 2721–2731.
- Kaulfers, A.M., Deka, R., Dolan, L., Martin, L.J., 2015. Association of INSIG2 polymorphism with overweight and LDL in children. *PLoS One* 10, e0116340.
- Lee, Y.H., Kim, W., Yu, B.C., Park, B.L., Kim, L.H., Shin, H.D., 2008. Association of the ins/del polymorphisms of uncoupling protein 2 (UCP2) with BMI in a Korean population. *Biochem. Biophys. Res. Commun.* 371, 767–771.
- Lindgren, C.M., Heid, I.M., Randall, J.C., Lamina, C., Steinthorsdottir, V., Qi, L., Speliotes, E.K., et al., 2009. Genome-wide association scan meta-analysis identifies three loci influencing adiposity and fat distribution. *PLoS Genet.* 5, e1000508.
- Marks, D.L., Boucher, N., Lanouette, C.M., Pérusse, L., Brookhart, G., Comuzzie, A.G., et al., 2004. Ala67Thr polymorphism in the Agouti-related peptide gene is associated with inherited leanness in humans. *Am. J. Med. Genet.* 126, 267–271.
- Martinez-Hervas, S., Mansego, M.L., de Marco, G., Martinez, F., Alonso, M.P., Morcillo, S., et al., 2012. Polymorphisms of the UCP2 gene are associated with body fat distribution and risk of abdominal obesity in Spanish population. *Eur. J. Clin. Invest.* 42, 171–178.
- Mitchell, T.M., 2006. The discipline of machine learning. *Mach. Learn.* 17, 1–7.
- Modrego, J., Azcona, L., Martín-Palacios, N., Zamorano-León, J.J., Segura, A., Rodríguez, P., et al., 2013. Platelet content of nitric oxide synthase 3 phosphorylated at Serine 1177 is associated with the functional response of platelets to aspirin. *PLoS One* 8, e82574.
- 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, 1523–1529.
- Nicoletti, C.F., de Oliveira, A.P., Brochado, M.J., Pinhel, M.A., de Oliveira, B.A., Marchini, J.S., et al., 2017. The Ala55Val and -866G > A polymorphisms of the UCP2 gene could be biomarkers for weight loss in patients who had Roux-en-Y gastric bypass. *Nutrition* 33, 326–330.
- Nigatu, Y.T., van de Ven, H.A., van der Klink, J.J., Brouwer, S., Reijneveld, S.A., Bültmann, U., 2016. Overweight, obesity and work functioning: the role of working-time arrangements. *Appl. Ergon.* 52, 128–134.
- Ollmann, M.M., Wilson, B.D., Yang, Y.K., Kerns, J.A., Chen, Y., Gantz, I., et al., 1997. Antagonism of central melanocortin receptors in vitro and in vivo by agouti-related protein. *Science* 278, 135–138.
- Orkunoglu-Suer, F.E., Gordish-Dressman, H., Clarkson, P.M., Thompson, P.D., Angelopoulos, T.J., Gordon, P.M., et al., 2008. INSIG2 gene polymorphism is associated with increased subcutaneous fat in women and poor response to resistance training in men. *BMC Med. Genet.* 9, 117.
- Podgorelec, V., Kokol, P., Stiglic, B., Rozman, I., 2002. Decision trees: an overview and their use in medicine. *J. Med. Syst.* 26, 445–463.
- Price, R.A., Li, W.D., Zhao, H., 2008. FTO gene SNPs associated with extreme obesity in cases, controls and extremely discordant sister pairs. *BMC Med. Genet.* 9, 4.
- R Development Core, 2011. Team: A Language and Environment for Statistical Computing. The R Foundation for Statistical Computing, Vienna, Austria 3-900051-07-0.
- Romieu, I., Dossus, L., Barquera, S., Blottière, H.M., Franks, P.W., Gunter, M., et al., 2017. Energy balance and obesity: what are the main drivers? *Cancer Causes Control* 28, 247–258.
- Rosmond, R., Bouchard, C., Björntorp, P., 2002. Lack of association between the uncoupling protein-2 Ala55Val gene polymorphism and phenotypic features of the metabolic syndrome. *Biochim. Biophys. Acta* 1588, 103–105.
- Ruiz, J.R., Larrarte, E., Margareto, J., Ares, R., Labayen, I., 2011. Role of β-adrenergic receptor polymorphisms on body weight and body composition response to energy restriction in obese women: preliminary results. *Obesity (Silver Spring)* 19, 212–215.
- Russell, S., Norvig, P., 2013. Artificial Intelligence: A Modern Approach, 3rd edn. Pearson, Harlow.
- Segal, E., Fondufe-Mittendorf, Y., Chen, L., Thåström, A., Field, Y., Moore, I.K., et al., 2006. A genomic code for nucleosome positioning. *Nature* 44, 772–778.
- Segura Fragoso, A., Rius Mery, G., 1999. Cardiovascular risk factors in a rural population of Castilla-La Mancha. *Rev. Esp. Cardiol.* 52, 577–588.
- Shutter, J.R., Graham, M., Kinsey, A.C., Scully, S., Lüthy, R., Stark, K.L., 1997. Hypothalamic expression of ART, a novel gene related to agouti, is up-regulated in obese and diabetic mutant mice. *Genes Dev.* 11, 593–602.
- Song, H.R., Park, H.S., Yun, K.E., Cho, S.H., Choi, E.Y., Lee, S.Y., et al., 2010. Gender and age differences in the impact of overweight on obesity-related quality of life among Korean adults. *Obes. Res. Clin. Pract.* 4, e1–e82.
- Song, Q.Y., Song, J.Y., Wang, Y., Wang, S., Yang, Y.D., Meng, X.R., et al., 2017. Association study of three gene polymorphisms recently identified by a genome-wide association study with obesity-related phenotypes in Chinese children. *Obes. Facts* 10, 179–190.
- Souren, N.Y., Paulussen, A.D., Steyls, A., Loos, R.J., Stassen, A.P., Gielen, M., et al., 2008. Common SNPs in LEP and LEPR associated with birth weight and type 2 diabetes-related metabolic risk factors in twins. *Int. J. Obes.* 32, 1233–1239.
- Tabara, Y., Kawamoto, R., Osawa, H., Nakura, J., Makino, H., Miki, T., et al., 2008. No association between INSIG2 Gene rs7566605 polymorphism and being overweight in Japanese population. *Obesity* 16, 211–215.
- Takeuchi, S., Katoh, T., Yamauchi, T., Kuroda, Y., 2012. ADRB3 polymorphism associated with BMI gain in Japanese men. *Exp. Diabetes Res.* 2012, 973561.
- Tremmel, M., Gerdtham, U.-G., Nilsson, P.M., Saha, S., 2017. Economic burden of obesity: a systematic literature review. *Int. J. Environ. Res. Public Health* 14, 435.
- Vink, T., Hinney, A., van Elburg, A.A., van Goozen, S.H., Sandkuijl, L.A., Sinke, R.J., et al., 2001. Association between an agouti-related protein gene polymorphism and anorexia nervosa. *Mol. Psychiatry* 6, 325–328.
- Willer, C.J., Speliotes, E.K., Loos, R.J., Li, S., Lindgren, C.M., Heid, I.M., et al., 2009. Six new loci associated with body mass index highlight a neuronal influence on body weight regulation. *Nat. Genet.* 41, 25–34.
- Xiu, L.L., Weng, J.P., Sui, Y., Wang, J., Yan, J.H., Huang, Z.M., et al., 2004. Common variants in beta 3-adrenergic-receptor and uncoupling protein-2 genes are associated with type 2 diabetes and obesity. *Zhonghua Yi Xue Za Zhi* 84, 375–379.
- Yasukochi, Y., Sakuma, J., Takeuchi, I., Kato, K., Oguri, M., Fujimaki, T., et al., 2018. Identification of three genetic variants as novel susceptibility loci for body mass index in a Japanese population. *Physiol. Genomics* 50, 179–189.
- Yu, X., Jacobs Jr., D.R., Schreiner, P.J., Gross, M.D., Steffes, M.W., Fornage, M., 2005. The uncoupling protein 2 Ala55Val polymorphism is associated with diabetes mellitus: the CARDIA study. *Clin. Chem.* 51, 1451–1456.
- Zamorano-León, J.J., Yañez, R., Jaime, G., Rodríguez-Sierra, P., Calatrava-Ledrado, L., Alvarez-Granada, R.R., et al., 2012. KCNH2 gene mutation: a potential link between epilepsy and long QT-2 syndrome. *J. Neurogenet.* 26, 382–386.
- Zhang, M., Wang, M., Zhao, Z.T., 2014. Uncoupling protein 2 gene polymorphisms in association with overweight and obesity susceptibility: a meta-analysis. *Meta Gene* 2, 143–159.

