

## Resumen

El cáncer de mama es uno de los tumores más frecuentes en la población femenina, estableciéndose una prevalencia global en los países occidentales alrededor del 12,5%, es decir, aproximadamente una de cada ocho mujeres padecerá cáncer de mama a lo largo de su vida.

En la lucha por la reducción del impacto de esta enfermedad en la sociedad y en las entidades sanitarias públicas, en los últimos años se han desarrollado diferentes propuestas, siendo la estratificación de la población general uno de dichos métodos. Estos sistemas de estratificación permiten clasificar a la población general en función del riesgo a padecer cáncer de mama de manera individual usando diferentes variables genéticas, fenotípicas o ambientales.

En la presente tesis, partiendo de una cohorte caso – control retrospectivo, basado en población española femenina, se aplicarán diferentes modelos matemáticos utilizando variables genéticas, fenotípicas y combinando ambas. Estos modelos permitirán la estratificación de la población general femenina en función del riesgo individual a padecer cáncer de mama esporádico.

En un primer estudio se evaluaron las variables genéticas como factor discriminante en la cohorte caso – control estudiado. Estas variables son variantes genéticas puntuales denominados SNPs (*single nucleotide polymorphism*). Estas variantes fueron seleccionadas mediante estudios GWAs (*genome-wide association study*) basados en cohortes casos-control, asociados a cáncer de mama esporádico y usando miles de mujeres de población general Caucásica. Estas alteraciones, estadísticamente significativas y asociadas a fenotipo, tienen escaso efecto funcional, poca penetrancia, son relativamente frecuentes en la población y con asociación individual a un riesgo bajo o moderado a padecer cáncer de mama.

En la presente tesis, se utilizaron inicialmente 121 de estos SNPs descritos anteriormente y combinados mediante modelos “*polygenic risk score*” (PRS) para evaluar su poder discriminante en función del riesgo individual a padecer cáncer de mama esporádico en una cohorte caso-control representativa de la población española.

En un segundo estudio se evaluaron las variables fenotípicas como factores discriminantes de la cohorte caso- control estudiado. Los factores de riesgo fenotípicos seleccionados se obtuvieron a partir de bibliografía científica mayoritariamente de población española y son; la densidad mamaria, el estatus de menopausia, la edad de la mujer, la edad de menarquia, edad el primer hijo nacido vivo y los posibles antecedentes familiares.

En un tercer estudio se evaluó la significancia y el poder discriminante utilizando un modelo logístico que combina las variables genéticas agrupadas mediante un PRS, las variables fenotípicas, y la interacción de algunas de estas variables. Se evaluó la posible utilidad del modelo para la clasificación de la cohorte representativa de la población española en función del riesgo individual a padecer cáncer de mama esporádico.

En la presente tesis se presenta un modelo matemático que combina factores genéticos y fenotípicos de riesgo a padecer cáncer de mama esporádico. En este estudio se demuestra la sinergia y la ausencia de colinealidad entre este tipo de variables, permitiendo obtener una capacidad discriminante entre las mujeres

que desarrollaron cáncer de mama esporádico durante los siguientes cinco años después de la toma de los datos y mujeres sanas que no lo desarrollaron. Esta capacidad discriminante obtenida mediante estos modelos combinados de los factores de riesgo es mayor que la obtenida usando modelos basados solo en los factores de riesgo individuales. También permite, como prueba de concepto, valorar su posible utilidad en un sistema de cribado mediante la clasificación de la población general en diferentes categorías en función de su riesgo a padecer cáncer de mama esporádico.