Expresión del gen TP53 en el tejido mamario de pacientes y de personas en riesgo de cáncer de pecho: análisis por grupos etarios

Proyecto de Herramientas para Ciencia de Datos I CA0204

Alexandra González | Estefanía Mora | Liz Salazar | José Miguel Rodríguez

Introducción

El gen TP53 es un supresor tumoral esencial involucrado en procesos como la regulación génica, el control del ciclo celular, la apoptosis y la estabilidad del genoma, según López et al. (2001). Además, participa en el mantenimiento de la integridad del ADN y en la supervivencia celular ante agentes dañinos. Cuando este gen se altera, aumenta significativamente el riesgo de cáncer: si la proteína p53 está mutada, el ciclo celular no se detiene pese al daño en el ADN, lo que provoca proliferación celular descontrolada y la formación de tumores. Patiño et al. (2004) señalan que TP53 está alterado en cerca del 60 % de los tumores y que alrededor de 2.4 millones de los casos de cáncer anuales se deben a mutaciones en este gen.

Con base en lo anterior, se propone estudiar su expresión en el tejido mamario, dado que el cáncer de mama "es el tipo de cáncer más frecuente y la causa más común de muerte por cáncer en mujeres" (Organización Panamericana de la Salud, s.f.). El análisis se realizará por grupos etarios: adolescencia (12–18), juventud (19–30), adultez (30–60) y vejez (60+).

Delimitación del tema

El proyecto se centra en la relación entre el gen TP53 y el cáncer de mama, analizando cómo la presencia o alteración del primero puede influir en el desarrollo del segundo. El estudio se organiza según los grupos etarios previamente definidos.

Primero, se examinarán los niveles de expresión del gen TP53 en tejido mamario de pacientes y personas en riesgo, utilizando una base de datos internacional. Luego, las conclusiones obtenidas se aplicarán al contexto costarricense mediante datos del Ministerio de Salud sobre casos de cáncer de mama, con el fin de plantear hipótesis sobre la posible relación entre la expresión de TP53 y la incidencia de esta enfermedad en Costa Rica.

Justificación

El análisis del gen TP53 es esencial para comprender los mecanismos moleculares del cáncer de mama, ya que sus mutaciones pueden generar proliferación celular descontrolada y aumentar el

riesgo de malignidad. Evaluar su expresión según grupos etarios permite identificar posibles patrones vinculados a factores biológicos o ambientales.

Asimismo, contrastar estos resultados con los datos del Ministerio de Salud de Costa Rica permite contextualizar los hallazgos en el ámbito nacional y apoyar el desarrollo de estrategias más efectivas de prevención y detección temprana.

Objetivos

Objetivo general

Analizar la expresión y frecuencia de las mutaciones del gen TP53 en el tejido mamario de pacientes y personas en riesgo de desarrollar cáncer de pecho, diferenciando los resultados por grupos etarios.

Objetivos específicos

- Identificar las mutaciones más frecuentes del gen TP53 presentes en el tejido mamario de los individuos analizados.
- Describir la distribución etaria de los casos diagnosticados con mutaciones en el gen TP53.
- Determinar la relación existente entre los tipos de mutación y la edad al diagnóstico.

Hipótesis

Hipótesis general:

La expresión del gen TP53 en el tejido mamario presenta variacione germinales del gen TP53 son más frecuentes en mujeres diagnosticadas con cáncer de mama a edades tempranas, especialmente menores de 40 años (Evans et al., 2020; Li et al., 2025; Kwong et al., 2020). Además, se ha reportado que alrededor del 30 % de los cánceres de mama presentan mutaciones en este gen, lo que respalda la importancia de analizarlo en cohortes estratificadas por edad (Silwal-Pandit et al., 2014; Hwang et al., 2024).

Hipótesis específicas:

s significativas según el grupo etario, siendo más frecuentes las mutaciones en adultos jóvenes y adultos medios.

Diversos estudios han mostrado que las mutaciones somáticas y - H_1: Las mutaciones más frecuentes del gen TP53 se concentran en individuos diagnosticados entre los 30 y 50 años.

Esta hipótesis se fundamenta en el hecho de que varias investigaciones han reportado picos de incidencia y prevalencia de mutaciones TP53 en grupos etarios jóvenes y de mediana edad. Por ejemplo, Li et al. (2025) encontraron prevalencias mayores en pacientes de 30 años (menores o

iguales) que en grupos de 31–40 años, y Kwong et al. (2020) reportaron edades de diagnóstico cercanas a los 27–30 años en portadoras de mutaciones germinales. Aunque los rangos exactos varían entre poblaciones, la evidencia coincide en una mayor frecuencia de mutaciones TP53 en edades anteriores a los 50 años, lo cual respalda la formulación de esta hipótesis.

- H_2: Existe una relación significativa entre el tipo de mutación (MUT_ID) y la edad al diagnóstico (Age_at_diagnosis).

Esto coincide con hallazgos que señalan que el espectro mutacional de TP53 no es uniforme y puede asociarse con factores clínicos, incluyendo edad, subtipo tumoral y agresividad (Silwal-Pandit et al., 2014). Además, Hwang et al. (2024) destacan que la distribución de mutaciones TP53 en cáncer de mama presenta patrones clínicos definidos, lo cual sugiere que podrían encontrarse relaciones estadísticamente significativas entre los tipos de mutación (hotspots, missense, truncantes, etc.) y la edad en que se diagnostica el tumor.

- H_3: Los rangos etarios del escenario costarricense que poseen una mayor acumulación de casos, también presentan una mayor diversidad mutacional del gen TP53.

Aunque no existen estudios extensivos de TP53 en poblaciones costarricenses, las tendencias globales muestran que poblaciones con mayor carga de cáncer de mama en ciertos grupos etarios tienden también a presentar variabilidad mutacional significativa en genes clave como TP53 (Evans et al., 2020; Silwal-Pandit et al., 2014). Por lo tanto, extrapolando estas tendencias, es razonable plantear que si en Costa Rica existe un grupo etario con mayor prevalencia de casos, podría observarse también una mayor diversidad mutacional asociada al gen.

Estas hipótesis se generaron de acuerdo a lo estipulado por la evidencia científica existente sobre la distribución etaria de mutaciones del gen TP53 en cáncer de mama, así como por estudios que destacan su variabilidad mutacional y su relevancia clínica en poblaciones jóvenes y de mediana edad (Evans et al., 2020; Li et al., 2025; Silwal-Pandit et al., 2014).

Marco metodológico

El estudio se desarrollará con un enfoque cuantitativo descriptivo, cuyo objetivo es analizar cómo varían la expresión y la frecuencia de las mutaciones del gen TP53 en tejido mamario según grupos etarios, sin manipular las variables, centrándose en su descripción, comparación y correlación (Hernández-Sampieri & Mendoza, 2018).

La investigación utilizará bases de datos secundarias, principalmente *The TP53 Dataset* (versión R21), que reúne información sobre miles de variantes y casos con mutaciones confirmadas (Bouaoun et al., 2016), seleccionando únicamente los datos correspondientes al tejido mamario. De forma complementaria, se emplearán los registros del Ministerio de Salud de Costa Rica (2022) para contrastar los hallazgos internacionales con la incidencia local del cáncer de mama.

En una primera fase se aplicarán estadísticas descriptivas —medidas de tendencia central y dispersión, tablas de frecuencia, histogramas y diagramas de caja— para caracterizar la población y detectar patrones preliminares en la distribución de mutaciones según la edad (Ott & Longnecker, 2016).

Luego se realizará una comparación entre grupos mediante ANOVA, siempre que los datos cumplan los supuestos requeridos (Montgomery, 2017); en caso contrario, se utilizará la prueba no paramétrica de Kruskal–Wallis (Gibbons & Chakraborti, 2011). Este análisis permitirá determinar si existen diferencias significativas en la edad al diagnóstico según el tipo de mutación del gen TP53.

También se efectuarán análisis de correlación entre la edad al diagnóstico (Age_at_diagnosis) y el número de casos reportados por consorcios como TCGA, ICGC y GENIE, aplicando correlaciones de Pearson o Spearman según la distribución de los datos (Benesty et al., 2009). Con ello se busca establecer si la edad se relaciona con la frecuencia o severidad de las mutaciones.

Asimismo, se analizarán frecuencias y proporciones para identificar la incidencia relativa de cada tipo de mutación por grupo etario, lo cual permitirá reconocer patrones de riesgo relevantes (Agresti, 2018).

Previo a los análisis inferenciales, se realizará una limpieza y depuración de datos, incluyendo el tratamiento de valores faltantes y la detección de atípicos mediante el IQR, para asegurar la calidad del conjunto de datos (Field, 2018).

Cada técnica aportará información complementaria: la estadística descriptiva mostrará la distribución de edades y mutaciones; ANOVA/Kruskal–Wallis revelarán diferencias significativas entre grupos; las correlaciones permitirán evaluar la relación entre edad y frecuencia de mutaciones; y las frecuencias identificarán los grupos con mayor incidencia.

Finalmente, estos patrones se contrastarán cualitativamente con los datos del Ministerio de Salud para contextualizar los resultados globales en Costa Rica y formular hipótesis sobre la posible prevalencia local, sin establecer correlaciones estadísticas directas por las diferencias entre ambas bases.

Bases de datos

La principal fuente de información será *The TP53 Dataset*, versión R21, desarrollada por el Instituto Nacional de Cáncer de Estados Unidos y actualizada en enero de 2025. Esta base contiene cerca de 29,900 variantes tumorales, más de 2,155 individuos con mutaciones confirmadas en *TP53* y datos funcionales sobre más de 9,000 proteínas mutantes. Para este estudio, se filtrará únicamente la información cuya morfología corresponda al tejido mamario.

En el caso de Costa Rica, se utilizará la base del Ministerio de Salud (2022) titulada Incidencia de tumores malignos: diferentes características, la cual segmenta los casos por sexo y grupos etarios. Su función será principalmente cualitativa, orientada a formular hipótesis sobre la relación entre TP53 y el cáncer de mama en la población nacional.

El análisis se fundamentará en la tabla derivada de la base internacional, tras un proceso de filtrado por relevancia de variables y por topografía igual a tejido mamario. Este procedimiento

generó un conjunto de 1,153 pacientes, mayoritariamente mujeres, con edades entre 14 y 90 años y de diversas nacionalidades. Dado que esta base se actualiza con información proveniente de literatura médica y de otros repositorios públicos, constituye la fuente cuantitativa central del proyecto, mientras que los datos nacionales servirán únicamente como complemento contextual.

Variables de estudio

El estudio analiza 1,153 individuos de entre 14 y 90 años, provenientes de distintas regiones y registrados por el Instituto Nacional del Cáncer de EE. UU. en enero de 2025. Todos son pacientes o personas con alto riesgo de cáncer de mama y muestran expresión del gen TP53 en tejido mamario. Cada individuo constituye una unidad de análisis. Se consideran variables demográficas (país, región, sexo, edad al diagnóstico), identificadores individuales y características clínicas del tumor (topografía, morfología y estado de vida o muerte).

Discusión y análisis de resultados

Estadísticas Descriptivas de la Edad

Edad promedio: 34 años. Edad mediana: 33 años. Edad máxima: 94 años. Edad mínima: 1 año

Primer cuartil: 19 años. Segundo cuartil: 33 años. Tercer cuartil: 47 años.

Desviación estándar: 19.2 años. Varianza: 369 años IQR: 28 años

Relación de edad al diagnóstico y mutaciones más frecuentes

Al analizar la relación entre la variable Age_at_diagnosis y MUT_ID usando las 10 mutaciones más frecuentes, se utilizó shapiro.test para comprobar si las variables se dstribuyen normalmente. Debido a que se presentaron pocos datos que si cumplen con esta condición, se escogió Kruskal-Wallis para determinar si hay una diferencia significativa entre los distintos grupos de mutaciones. Esto se observa mediante la Figura 1.

Evaluación del método Kruskal-Wallis para encontrar diferencias significativas entre edades al diagnóstico de distintos grupos.

Mediante la aplicación de Kruskal-Wallis se detectó un valor p menor que 0,05 lo que significa que entre algunos de los grupos analizados hay una diferencia significativa respecto a las edades al diagnóstico. A partir de ellos, se aplicó dunnTest que ayuda a separar pares de grupos para que diferencias hay entre ellos y así detectar los tipos de mutaciones con mayor o menor edad de diagnóstico.

Diferencia de rangos y comparación entre mutaciones

En la Figura 2 se presentan diferencias de rangos promedio entre pares de mutaciones. Cada punto es una comparación y los colores son determindados de acuerdos a si el valor p es menor a 0,05 (rojos) y o sino (grises), es decir, si se presenta una diferencia estadística significativa.

El eje vertical permite observar que mutación presenta un diagnóstico mayor respecto a la mutación con la que se compara, si la diferencia es positiva, el primera mutación del par es la que tiende a asociarse con una edad de diagnóstico tardío. Si la diferencia es negativa, el primer par es, en cambio, el que se relaciona con una edad temprana de diagnóstico.

Por lo tanto, las mutaciones 4585 y 2242 exponen una tendencia a asociarse con edad más altas de diagnóstico, al compararse con las mutaciones 3294, 3297, 2705 y 2143. Además, se encontró que las mutaciones 2705, 3297 y 3294, se asocian a edades más tempranas al compararse con 2242.

Distribución de edades y edades al diagnóstico

El histograma (Figura 3) muestra que la distribución de la edad al diagnóstico es asimétrica hacia la derecha (sesgo positivo). La mayoría de los casos se concentran entre los 25 y 40 años, con un pico marcado alrededor de los 30 años, lo que indica que este grupo etario es el más común en los registros disponibles.

A partir de los 45–50 años la frecuencia comienza a disminuir de manera notable, y existe una cola extendida hacia edades más avanzadas (hasta alrededor de 75–80 años), pero con muy pocos casos. Esta forma confirma que la distribución no es normal y que la base contiene una gran proporción de diagnósticos en edades relativamente jóvenes.

El segundo histograma, Figura 4, aunque basado en un subconjunto más pequeño debido a la presencia de muchos valores NA (esto debido a la alta mortalidad que presentan los pacientes con cáncer), presenta una forma muy similar: un claro sesgo hacia la derecha y mayor concentración de casos entre los 25 y 40 años, con un máximo alrededor de los 30 años.

La menor cantidad total de observaciones hace que las barras sean más variables y que la forma no se vea tan suave, pero la tendencia general se mantiene: hay pocos casos en edades muy jóvenes (<20) y muy pocos casos por encima de los 60 años.

Ambos histogramas reflejan una tendencia consistente:

La mayoría de los diagnósticos se dan en personas entre 25 y 40 años. La distribución presenta una cola hacia la derecha, indicando sesgo positivo. No se observa una distribución normal. La presencia de NA afecta la claridad del segundo histograma, pero no modifica la tendencia central.

Estos patrones apoyan parcialmente la hipótesis de que las alteraciones del gen TP53 en tejido mamario tienden a concentrarse en grupos etarios relativamente jóvenes.

Evaluación de la normalidad de la correlación entre la edad y edad al diagnóstico con el conteo de casos de cáncer de acuerdo a consorcios internacionales.

Se evaluó la normalidad de las variables Edad (Age), Edad al diagnóstico (Age_at_diagnosis) y el conteo de casos (TCGA_ICGC_GENIE_count) mediante la prueba de Shapiro–Wilk. En los tres casos se obtuvo un valor de p extremadamente pequeño ($p < 0.05$), indicando que ninguna de las variables sigue una distribución normal:

$W_{\text{Age}} = 0.929,\ p = 1.3 \times 10^{-9}$

$W_{\text{Age\_at\_diagnosis}} = 0.916,\ p < 2.2 \times 10^{-16}$

$W_{\text{TCGA\_ICGC\_GENIE\_count}} = 0.653,\ p < 2.2 \times 10^{-16}$

Dado que las variables no cumplen el supuesto de normalidad se utilizó el coeficiente de correlación de Spearman (ρ) para analizar la relación entre la edad y el número de mutaciones reportadas por los consorcios TCGA, ICGC y GENIE.

En primer lugar, la correlación entre la edad (Age) y el conteo combinado de mutaciones (TCGA_ICGC_GENIE_count) no fue significativa: $\rho = 0.073,\ p = 0.246$. El valor de ρ, cercano a cero, indica que no existe una relación relevante entre la edad del paciente y la cantidad de mutaciones reportadas por los consorcios. Por otro lado, se evaluó la correlación entre la edad al diagnóstico (Age_at_diagnosis) y el mismo conteo. En este caso se obtuvo un coeficiente negativo muy pequeño: $\rho = -0.077,\ p = 0.011$. Aunque la asociación es estadísticamente significativa, la magnitud del coeficiente es extremadamente baja, lo que indica que la relación (aunque detectable debido al tamaño de la muestra) carece de relevancia práctica. En términos reales, la edad al diagnóstico apenas se relaciona con el número de mutaciones reportadas.

En conjunto, ambos análisis sugieren que ni la edad ni la edad al diagnóstico se asocian de manera consistente o relevante con el conteo de mutaciones TP53 provenientes de los consorcios TCGA, ICGC y GENIE.

Gráficamente, se puede observar mediante Figura 5 y Figura 6.

Relación entre la edad y edad al diagnóstico con el conteo de mutaciones reportadas por los consorcios internacionales

Los gráficos de dispersión (Figura 5 y 6) muestran visualmente la relación entre la edad y el conteo de mutaciones reportadas por los consorcios TCGA, ICGC y GENIE. En el caso de la variable Age, los puntos se encuentran ampliamente dispersos y la línea de tendencia lineal es casi horizontal, lo que confirma la ausencia de una relación clara entre ambas variables.

Para la variable Age_at_diagnosis, aunque la línea de tendencia presenta una leve pendiente negativa, la dispersión sigue siendo elevada, indicando que la relación es muy débil. Esto coincide con el valor de ρ obtenido en la correlación, que muestra una asociación estadísticamente significativa pero prácticamente irrelevante.

Incidencia del cáncer de mama en la población costarricense durante el año 2022

Nivel general

Durante el año 2022 la población costarricense presentó una totalidad de 1232 casos de cáncer de mama, lo anterior con una particularidad: a partir del gráfico de la distribución de casos según el sexo para Costa Rica (Figura 9), se determina la existencia de una brecha considerable en la incidencia del cáncer de mama en las personas de sexo masculino respecto a las personas de sexo femenino; los masculinos presentaron únicamente seis casos, mientras que las féminas acumularon una cantidad de 1226 casos. El fenómeno descrito no corresponde a un hecho aislado, tal como menciona Soto Flores (2015) " el cáncer de mama es la causa líder de muerte en mujeres de países en vías de desarrollo y la segunda causa de muerte en países desarrollados, siendo segundo al cáncer de pulmón" (p. 799). Asimismo, este autor se refiere al escenario nacional: "dentro de la población costarricense el cáncer de mama es la causa más común de mortalidad en mujeres por neoplasia maligna" (p. 800).

Del párrafo precedente se resalta un hecho importante, el cual consiste en que la mayoría de diagnósticos de cáncer de mama en el 2022 se presentaron en el sexo femenino, con un 99.51% del total.

Sexo masculino

Como se mencionó antes, de los 1232 casos solo seis fueron diagnosticados en sujetos de sexo masculino, equivalente a un 0.49% del total. En lo que respecta a la distribución de los casos, considerando el gráfico que describe dicho aspecto de acuerdo a los rangos etarios en el sexo masculino, se determina que la mayoría de casos se situaron en la categoría que comprende de los 65 a los 69 años, con una acumulación de dos elementos. Además, es importante destacar la inexistencia de diágnosticos de cáncer de mama en masculinos menores a 45 años, a la vez que se visualiza una concentración de los casos a partir de los 60 años. Se puede observar mediante la Figura 7.

Sexo femenino

En el caso del sexo femenino, se destaca la inexistencia de diagnósticos en personas menores a 20 años, lo cual representa un rango mucho menor en comparación al sexo masculino. Asimismo, la agrupación de edades que acumula la mayor cantidad de casos corresponde a aquella que comprende de los 55 a los 59 años con un total de 168. La variación de esta cifra respecto a otras agrupaciones es muy pequeña: el grupo a partir de los 75 años posee una cantidad de 161 elementos, de igual manera, el rango de edad que va de los 65 a los 69 años contiene 156 elementos.

Por otro lado, se visualiza una concentración de los casos a partir de los 40 años, donde las cifras acumuladas en los rangos etarios son mayores a 100. Previo a los 40 años, a excepción del rango que va de los 35 a los 39 años, todos acumulan cantidades menores a 20 elementos. En particular, se destaca que la menor cantidad de diagnósticos se sitúa en el rango etario que comprende de los 20 a los 24 años, con apenas dos elementos. Se puede observar mediante la Figura 8.

Relación del escenario costarricense en el 2022 y la expresión mutada del gen TP53

A partir de lo analizado en los subapartados anteriores, en conjunto con lo visualizado en los gráficos correspondientes a la incidencia del cáncer de mama en Costa Rica durante el 2022, se determina que en la población costarricense los tres rangos etarios con una mayor incidencia del cáncer estudiado fueron: de los 55 a los 59 años, a partir de los 75 años y de los 65 a los 69 años.

Se analizará la relación del boxplot que detalla la relación entre las diez principales mutaciones del gen TP53 y la edad al diagnóstico, con los tres rangos etarios de mayor incidencia en el escenario costarricense durante el año 2022.

Rango etario con mayor incidencia: de los 55 a los 59 años

En primera instancia, es preciso destacar que para todas las mutaciones se cumple que el 50% de los datos centrales se sitúan por debajo de los 50 años, además, se presentan casos particulares en los cuales los bigotes de las cajas están por debajo de esa misma edad o apenas la superan, tal es el escenario para las mutaciones de identificador 3297, 2143, 3294, 3737 y 3879. En particular, para la mutación con identificador 1429, se identifica la presencia de valores atípicos en edades que comprenden o son cercanas al rango etario de mayor incidencia en el país. Es importante aclarar que cuando se menciona que las mutaciones poseen valores atípicos en ciertas edades, se refiere a que la expresión de dicha mutación no es común para esas edades.

Por otra parte, la extensión de los bigotes de las mutaciones 2242 y 4585 comprenden el rango etario que va de los 55 a los 59 años: si bien, la mayoría de expresiones de ambas mutaciones no se sitúan en las edades versadas, su presencia en ellas no se considera un valor atípico e inclusive para la mutación 4585 representan las edades más altas consideradas como valores normales o típicos.

Rango etario con la segunda mayor incidencia: a partir de los 75 años

Para el rango etario bajo estudio, se presenta un fenómeno particular respecto a la expresión de las diez mutaciones del gen TP53 más comunes: en la totalidad de estas diez mutaciones, no se considera un valor típico su expresión en edades iguales o superiores a los 75 años. De hecho, en el gráfico no se muestra la presencia de valores atípicos en estas edades, a excepción de la mutación con identificador 4584, que posee un valor atípico en dicha área. Lo anterior no es sinónimo de que la expresión de las mutaciones sea común para las edades comprendidas en el rango etario, en su lugar sugiere que estas no se presentan del todo, ni siquiera como valor atípico.

Rango etario con la tercer mayor incidencia: de los 65 a los 69 años

Se identifican dos aspectos principales. En primer lugar, la presencia de valores atípicos situados en el rango etario bajo estudio para tres mutaciones distintas: 2143, 3294 y 3236. En segundo lugar, la extensión del bigote de la mutación de identificador 2242 comprende las edades

estudiadas; si bien, la mayoría de expresiones de la mutación 2242 no se sitúa en ellas, su presencia en tal rango etario no se considera un valor atípico.

Conclusiones

Primera hipótesis

- H_1: Las mutaciones más frecuentes del gen TP53 se concentran en individuos diagnosticados entre los 30 y 50 años.

Basados en los resultados obtenidos en el gráfico boxplot, se puede llegar a la conclusión de que el rango de edad concentrado entre los 55 y 20 años, con mayores frecuencias entre los 30 y 40 años. Por otro lado se obtuvo que la edad promedio fue de 34 años, mientras que la mediana de 33 años, lo que indica que no hay un sesgo a edades muy jóvenes o muy avanzadas; también se puede decir que, en general las mutaciones del gen TP53 no está en gran medida en pacientes de más de 50 años, sino que se está concentrando en edades de menos de 50.

Además, se obtuvo que la mutación más frecuente fue la del identificador 4585 con 314 observaciones, representando un 6.754% del total de observaciones, casi el doble que la mutación en el segundo lugar, la cual tiene el identificador 2143 con 180 observaciones, que representa un 3.872%, y la tercer mutación con mayor frecuencia fue el 3236 con 160 observaciones, representando un 3.442% del total de mutaciones.

Segunda hipótesis

- H_2: Existe una relación significativa entre el tipo de mutación (MUT_ID) y la edad al diagnóstico (Age_at_diagnosis).

El análisis mediante Kruskal–Wallis mostró que algunas mutaciones del gen TP53 sí presentan diferencias significativas en la edad al diagnóstico, evidenciando que la distribución por edad no es homogénea entre las mutaciones más frecuentes. Los análisis post-hoc identificaron mutaciones asociadas a diagnósticos más tardíos (4585 y 2242) y otras relacionadas con diagnósticos tempranos (2705, 3297 y 3294). Esto respalda parcialmente la hipótesis, ya que la relación solo aparece en ciertos tipos de mutación.

Por otro lado, el análisis correlacional con la variable sustituta TCGA_ICGC_GENIE_count mostró que ni la edad ni la edad al diagnóstico explican la cantidad de mutaciones reportadas por los consorcios internacionales. Aunque la correlación con la edad al diagnóstico fue estadísticamente significativa, su magnitud fue extremadamente débil ($\rho = -0.077$), por lo que carece de utilidad práctica.

Por lo tanto, con base en estos resultados, la hipótesis H2 no puede considerarse apoyada, al menos desde la perspectiva del análisis correlacional asignado. Los datos sugieren que la edad al diagnóstico no está asociada de manera relevante con la frecuencia de mutaciones observadas, por lo

que otros factores biológicos, genéticos o clínicos podrían ser más determinantes en la variabilidad mutacional.

Tercera hipótesis

- H_3: Los rangos etarios del escenario costarricense que poseen una mayor acumulación de casos, también presentan una mayor diversidad mutacional del gen TP53.

El análisis del escenario costarricense durante el año 2022 muestra que el sector etario con mayor incidencia de cáncer de mama, corresponde al rango que va de los 55 a los 59 años. Después, se encuentran las edades de más de 75 años y finalmente se encuentran las edades de 65 a 69 años. Además, es preciso destacar la existencia de una considerable brecha entre la cantidad de diagnósticos reportados en el sexo masculino respecto al sexo femenino: alrededor del 99.51% del total están asociados al sexo femenino.

Al comparar los resultados de Costa Rica durante el año 2022, con el comportamiento de expresión observado en las diez mutaciones más comunes del gen TP53, se identifican ciertas variaciones de acuerdo al rango etario bajo estudio, sin embargo, se determina que, de forma general que las expresiones mutacionales del gen TP53 no se presentan usualmente dentro de los rangos etarios en los cuales Costa Rica presenta la mayor incidencia. El análisis del boxplot, evidencia que para la totalidad de las diez mutaciones principales del gen, se cumple que el 50% de los datos centrales se sitúan por debajo de los 50 años, asimismo, son escasas las ocasiones en las que los bigotes alcanzan los rangos etarios de mayor incidencia para el escenario nacional durante el 2022. Así, la presencia de las mutaciones en los rangos especificados no es significativa y en su mayoría representa valores normales extremos o valores atípicos. En particular, en los grupos de 65 a 69 años y de 75 años en adelante, las mutaciones prácticamente tienden a no expresarse, a excepción de ciertos valores atípicos.

En conjunto, estos hallazgos sugieren que no existe una fuerte concordancia entre las agrupaciones etarias que acumularon la mayor incidencia del cáncer de mama en Costa Rica y la distribución de las edades asociadas a las mutaciones más comunes del gen TP53: las mutaciones tienden a expresarse con mayor frecuencia en adultos jóvenes y adultos de mediana edad, mientras que en el contexto costarricense los casos se concentran en edades más avanzadas.

Anexos

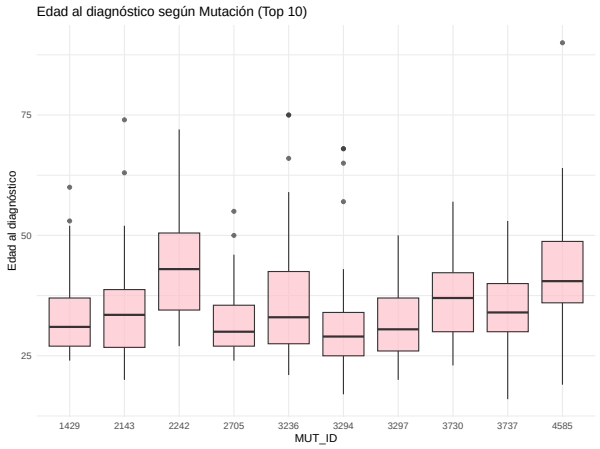

Figura 1. Relación de edad al diagnóstico y mutaciones más frecuentes

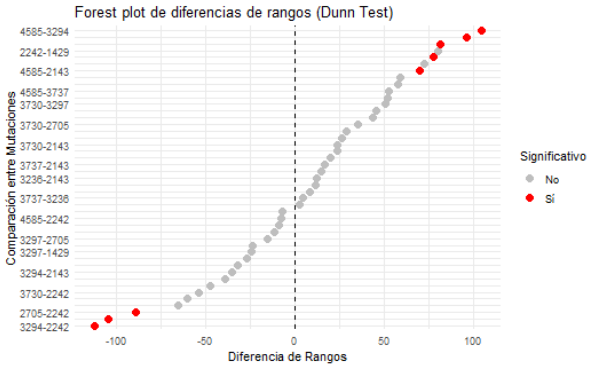

Figura 2. Diferencia de rangos y comparación entre mutaciones

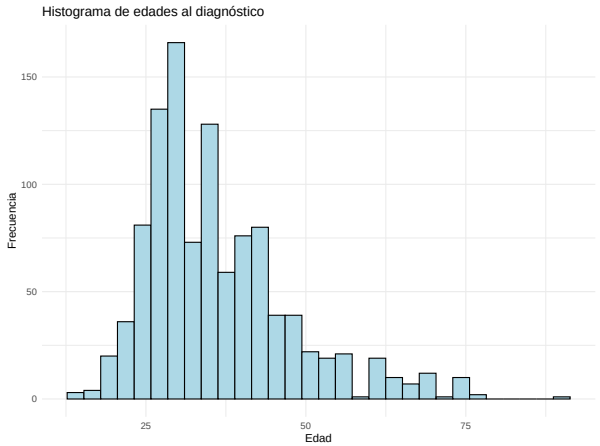

Figura 3. Histograma de edades al diagnóstico

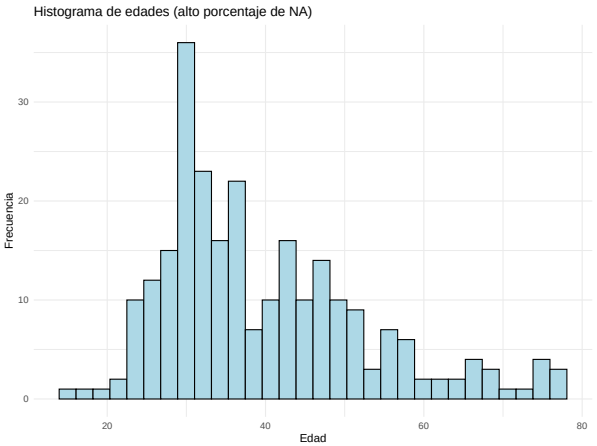

Figura 4. Histograma de edades

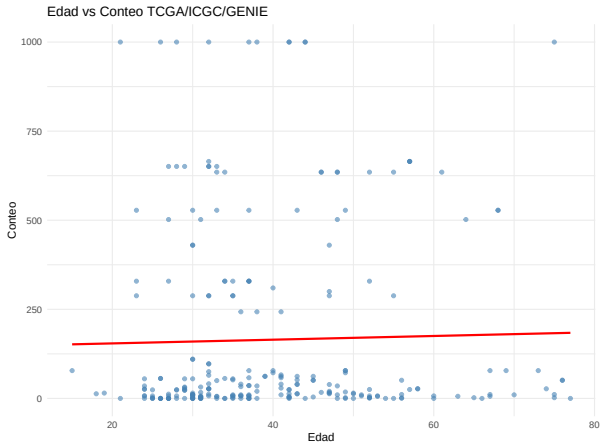

Figura 5. Correlación de edad con el conteo de mutaciones

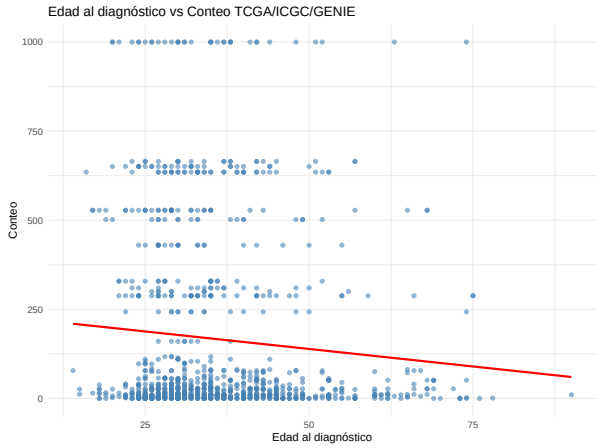

Figura 6. Correlación de edad al diagnóstico con el conteo de mutaciones

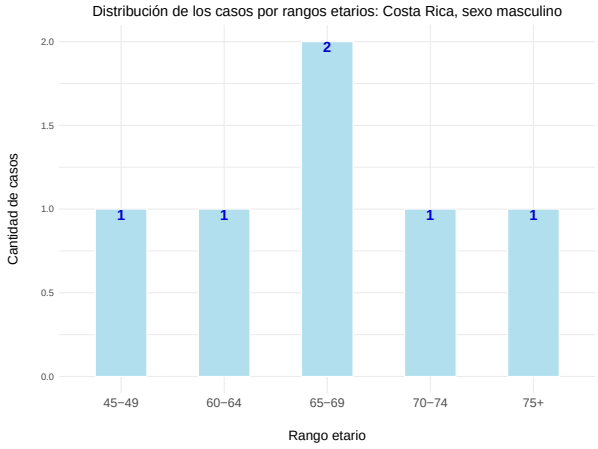

Figura 7. Distribución de casos CR hombres

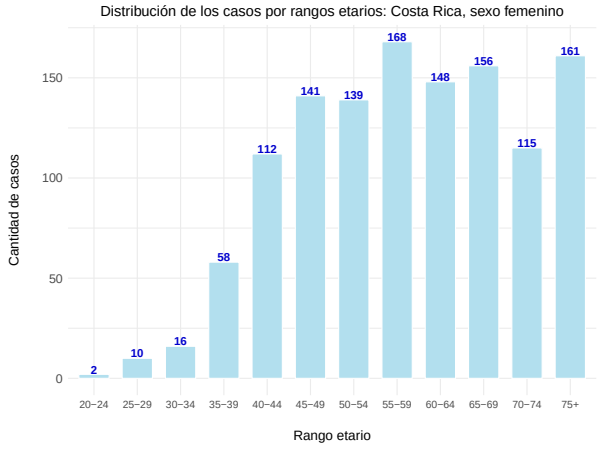

Figura 8. Distribución de casos CR mujeres

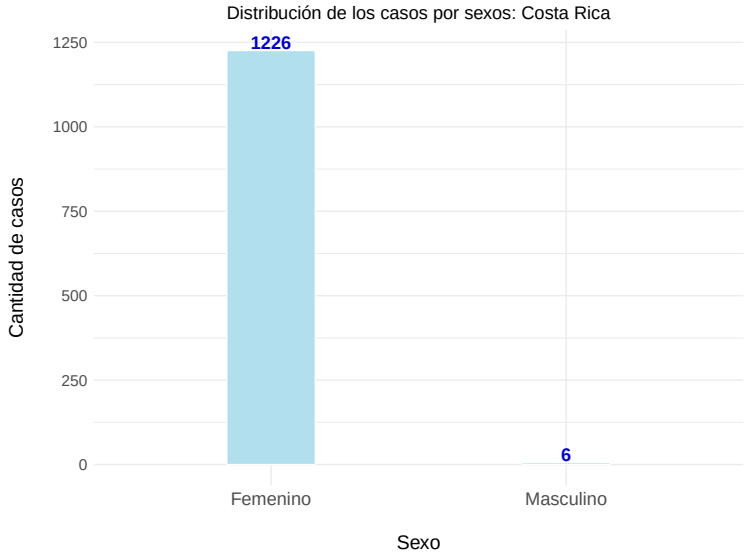

Figura 9. Distribución de los casos por sexos: Costa Rica

Referencias

Agresti, A. (2018). *Statistical Methods for the Social Sciences* (5th ed.). Pearson. https://www.pearson.com/en-us/subject-catalog/p/statistical-methods-for-the-social-sciences/P200000006673/9780134507101

Benesty, J., Chen, J., Huang, Y., & Cohen, I. (2009). Pearson Correlation Coefficient. En *Noise Reduction in Speech Processing* (pp. 1–4). Springer. https://doi.org/10.1007/978-3-642-00296-0_5

Bouaoun, L., Sonkin, D., Ardin, M., Hollstein, M., Byrnes, G., Zavadil, J., & Olivier, M. (2016). TP53 variations in human cancers: New lessons from the IARC TP53 database. *Human Mutation*, 37(9), 865–876. https://doi.org/10.1002/humu.23035

Field, A. (2018). *Discovering Statistics Using IBM SPSS Statistics* (5th ed.). Sage. https://uk.sagepub.com/en-gb/eur/discovering-statistics-using-ibm-spss-statistics/book257518

Gibbons, J. D., & Chakraborti, S. (2011). *Nonparametric Statistical Inference* (5th ed.). Chapman & Hall/CRC. https://doi.org/10.1201/b10905

Hernández-Sampieri, R., & Mendoza, C. (2018). *Metodología de la investigación: Las rutas cuantitativa, cualitativa y mixta.* McGraw-Hill. https://www.mheducation.com.mx/metodologia-de-la-investigacion-las-rutas-cuantitativa-cualitativa-y-mixta-9781456266669.html

Ministerio de Salud de Costa Rica. (2022). *Anuario de estadísticas vitales y de salud.* Ministerio de Salud de Costa Rica. https://www.ministeriodesalud.go.cr

Montgomery, D. C. (2017). *Design and Analysis of Experiments (9th ed.).* Wiley. https://www.wiley.com/en-us/Design+and+Analysis+of+Experiments%2C+9th+Edition-p-9781119113478

Ott, R. L., & Longnecker, M. (2016). *An Introduction to Statistical Methods and Data Analysis (7th ed.).* Cengage Learning. https://www.cengage.com/c/an-introduction-to-statistical-methods-and-data-analysis-7e-ott/

Soto Flores, W. (2015). Cáncer de mama. *Revista Médica de Costa Rica y Centroamérica*, (617), 799-802. https://www.binasss.sa.cr/revistas/rmcc/617/art20.pdf