

# 1 Avance 1 proyecto grupal

## **Integrantes del grupo:**

Alexandra González Bermúdez C4F535

Estefanía Mora Contreras C4H522

Liz Salazar Amaya C4J618

Jose Miguel Rodríguez Gómez C4J104

## **Tema escogido:**

Expresión del gen TP53 en tejido pulmonar humano para determinar susceptibilidad al desarrollo de cáncer de pulmón: efecto por edad y sexo en la población costarricense en el periodo 2017-2022

## **En qué consiste el proyecto:**

El proyecto consiste en analizar la expresión del gen TP53 en tejido pulmonar, mediante un análisis exploratorio computacional por medio de RStudio utilizando bases de datos públicas de transcriptómica que incluyen información de expresión génica, junto con las variables demográficas de los donantes de tejido. El objetivo del estudio es ver patrones de susceptibilidad (separados por sexo y edad) a enfermedades pulmonares relacionadas con la expresión del gen TP53. Además, se busca analizar como la edad y sexo influyen en la regulación de este gen y relacionar estos hallazgos en particular con el cáncer de pulmón, dado el papel central que desempeña TP53 como gen supresor tumoral. Todo esto se plantea dentro del contexto costarricense, y se utilizarán bases de datos, especialmente, del Ministerio de Salud para contrastar con otras bases de datos internacionales y medir mediante ciertas pruebas estadísticas la susceptibilidad de la población costarricense por sexo y edad a desarrollar cáncer de pulmón de acuerdo a la expresión del gen TP53. Se utilizarán las bases de datos del Ministerio de Salud de los años 2017 a 2022, los cuales son los más recientes.

## **Cómo se piensa aplicar R:**

Se plantea como base del proyecto el uso de RStudio para el manejo de datos, análisis estadístico y visualización de los resultados. Mediante el lenguaje R, se plantea importar los archivos de bases de datos públicas como GTEx o GEO, que esencialmente se tratan de tablas con valores de expresión génica incluyendo el gen TP53, y por medio de R se llevaría a cabo tanto la depuración de los datos como la comparación por edad y sexo, utilizando DESeq2 o edgeR, y también recount, GEOquery y TCGAbiolinks. Una vez organizada la información, se plantea el cálculo de diferencias de expresión del gen, lo cual se puede llevar a cabo por pruebas estadísticas sencillas como comparación de medias o análisis de regresión. Una vez obtenidos los resultados, se organizarán y visualizarán mediante las funciones básicas de R para representar los patrones encontrados. Esto incluye gráficos de cajas y de dispersión, y se puede utilizar ggplot2 para la elaboración de los gráficos, que permitirán la interpretación de los resultados del estudio.

## **Por qué se eligió:**

El cáncer de pulmón está entre las principales causas de mortalidad por cáncer en Costa Rica, especialmente en adultos mayores y en hombres. Al analizar cómo la edad y el sexo influyen en la expresión de TP53, que es un gen fundamental en la regulación del ciclo celular, se pueden generar hipótesis sobre susceptibilidad en la población costarricense a enfermedades como el cáncer de pulmón. Para esto, se utilizarán grandes repositorios internacionales (GTEx, GEO, Human Protein Atlas), y los resultados se contrastarán con estadísticas locales sobre incidencia y mortalidad de cáncer de pulmón en Costa Rica (datos del Registro Nacional de Tumores del Ministerio de Salud).

**Bases de datos:**

- GTEx (Genotype-Tissue Expression Project): metadatos de edad y sexo de donantes de tejidos. Datos de RNA-seq en tejidos pulmonares.
- TCGA (The Cancer Genome Atlas – Lung adenocarcinoma y squamous cell carcinoma): para contrastar tejidos normales/sanos de tejidos cancerosos.
- HPA (Human Protein Atlas): expresión de TP53 en diferentes tejidos. ARN y proteína.
- NCBI GEO (Gene Expression Omnibus): datasets de RNA-seq de pulmón divididos/ordenados por sexo y edad.
- Single Cell Expression Atlas (EMBL-EBI): expresión a nivel celular de tejido pulmonar.
- Ministerio de Salud (Costa Rica): <https://www.ministeriodesalud.go.cr/index.php/biblioteca-de-archivos-left/documentos-ministerio-de-salud/material-informativo/material-publicado/estadisticas-y-bases-de-datos/estadisticas-y-bases-de-datos-vigilancia-de-la-salud/estadistica-de-cancer-registro-nacional-tumores/incidencia/incidencia-2017-al-2022>