



Proyecto semestral

Retratos moleculares de tumores de
cancer de mama

Martinez Piza Erick
Ramirez Muñoz Sergio Ivan Alejandro
Tiscareño Galván Daniela

Índice

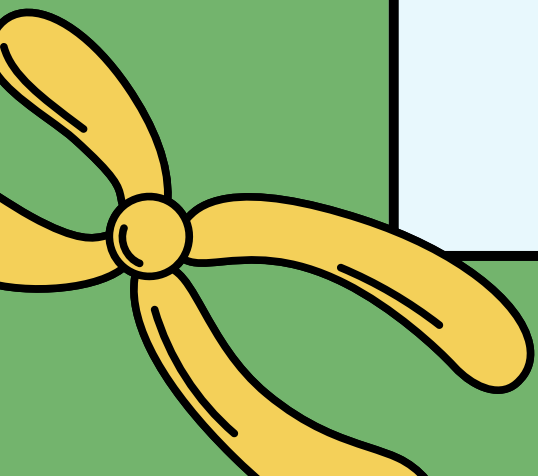
Introducción

Método

Conjunto de datos

Resultados

Bibliografía



Introducción

Receptores hormonales positivos, Luminal A



30-45% de las pacientes

Receptores hormonales negativos y receptores HER2



15-20 % de los tumores de mama

Receptores hormonales positivos y HER2 positivo o HER2 negativo, Luminal B



10-20% de los tumores de mama

Triple Negativos: sin receptores hormonales positivos ni receptores de HER2



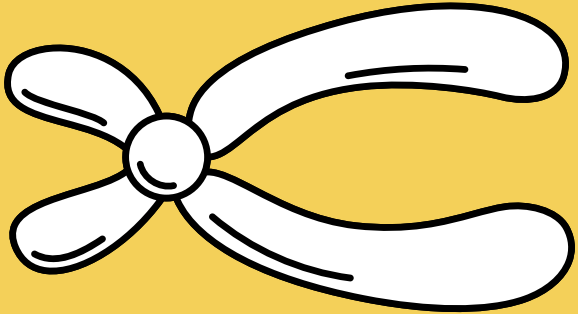
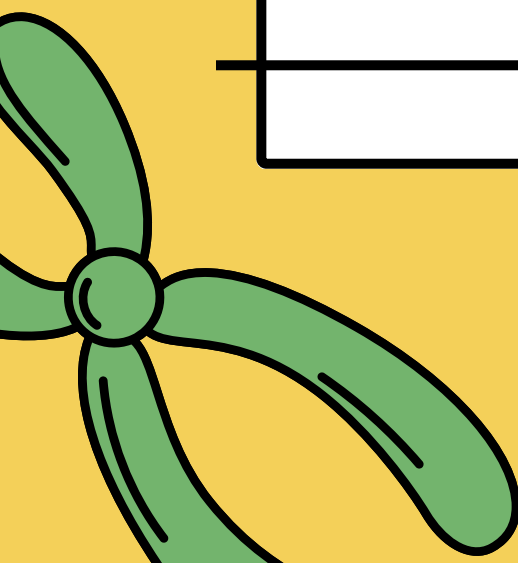
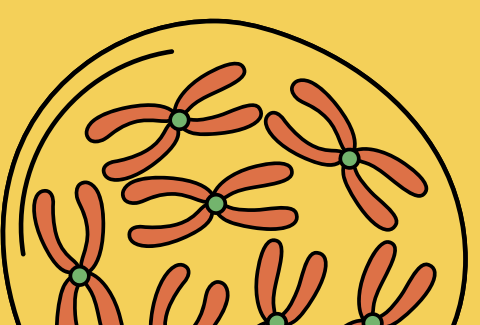
10%-15% de los tumores de mama

- 2.3 millones de mujeres fueron diagnosticadas con cáncer de mama
- Necesidad urgente de estrategias de gestión y tratamiento eficaces
- Cuatro subtipos de cáncer de mama



PREGUNTA DE INVESTIGACIÓN

¿Se puede crear un programa eficiente en python para poder identificar subtipos de cáncer recibiendo como entrada una secuencia de RNA?



Objetivos

Identificar subtipos de cáncer de mama a partir de una secuencia de scRNA

Identificar subtipos de cáncer de mama a partir de una secuencia de scRNA

Método conCluster

Paso 1: Filtración de genes

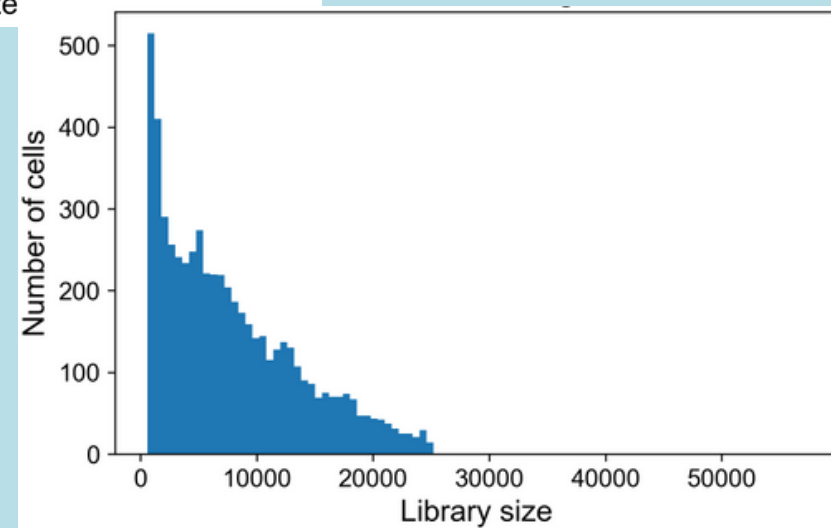
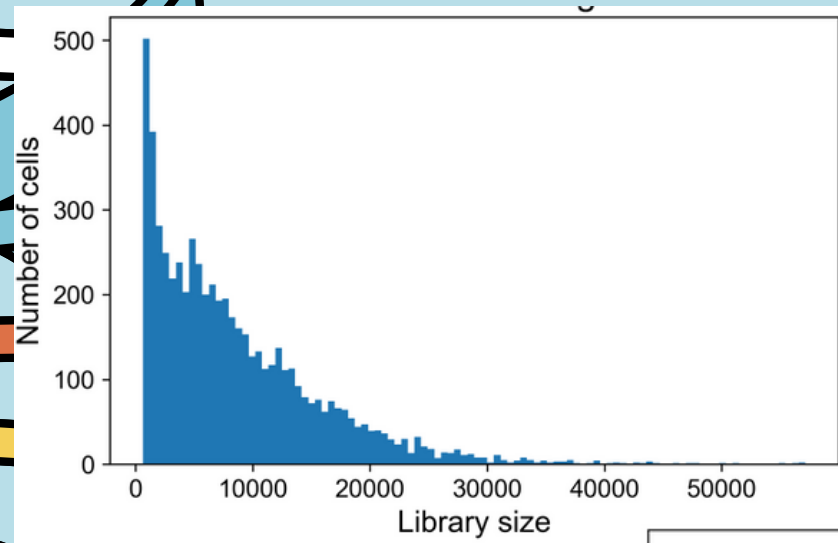
Paso 2: Reducimos dimensión usando t-SNE

Paso 3: Partición de células

Paso 4: Consenso de Clusters

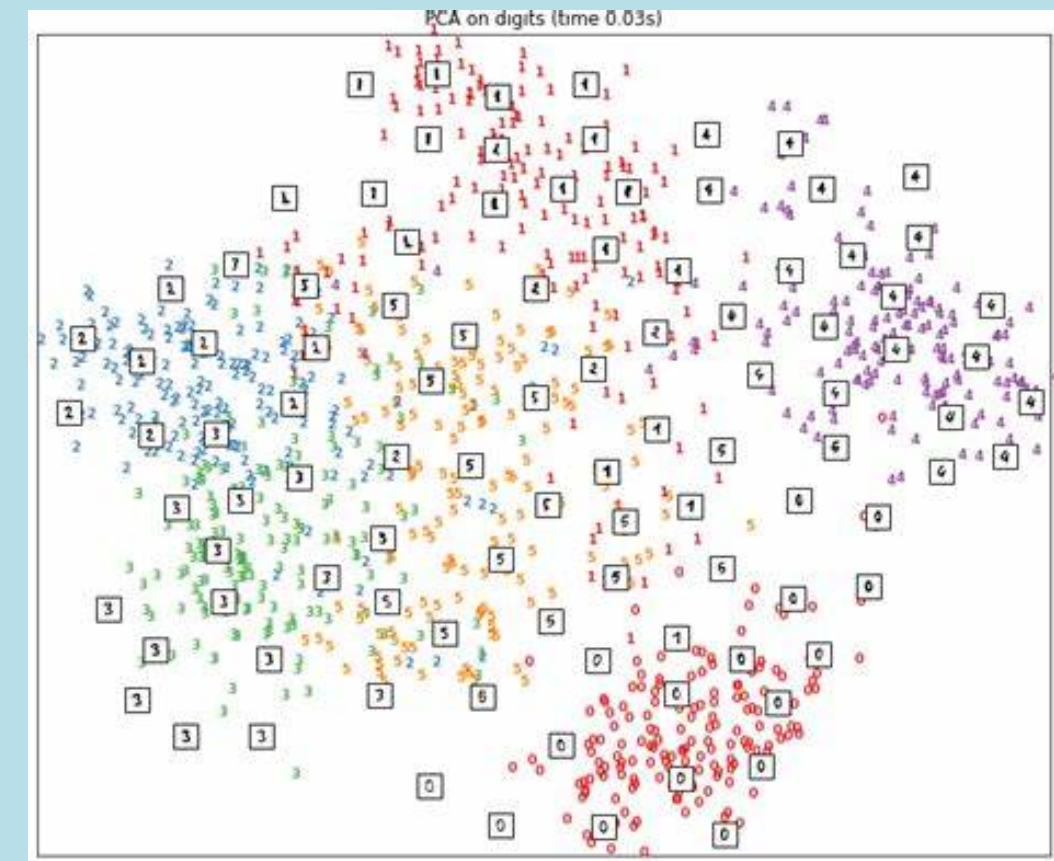
FILTRACIÓN DE GENES

- Se filtran los genes raros y ubicuos
- Identificación de genes más variables en el conjunto



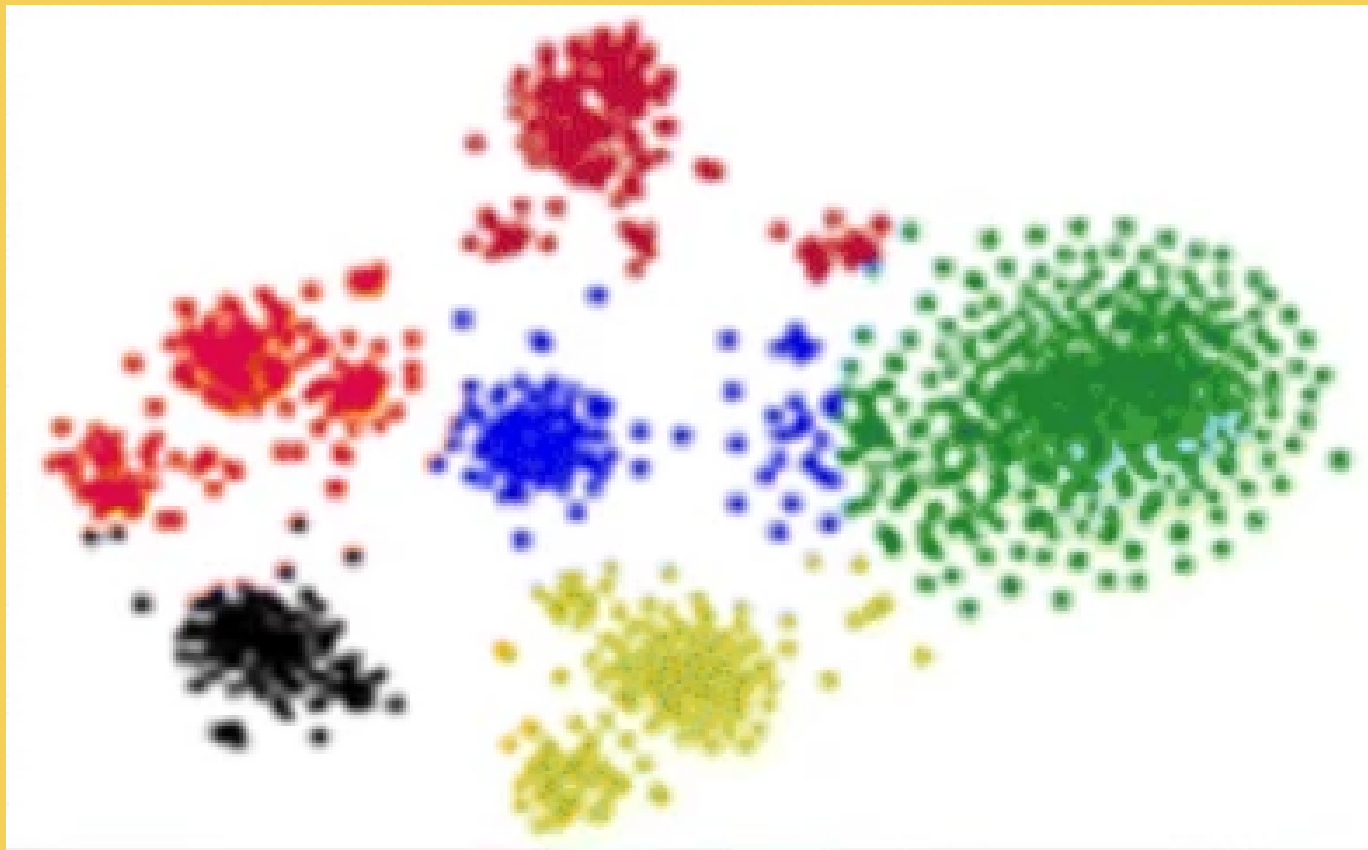
REDUCCIÓN DE DIMENSIÓN

- Se usa el método t-SNE
- Se convierte en un espacio dimensional de menor tamaño
- Parámetro de perplejidad



PARTICIÓN DE CÉLULAS

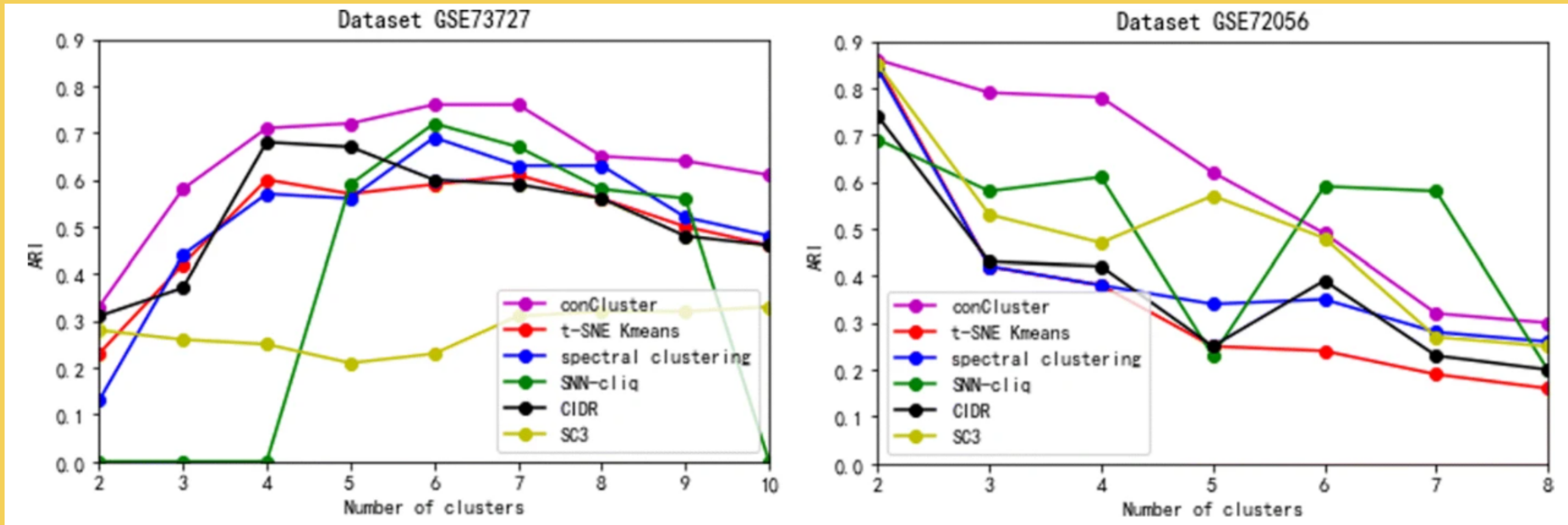
- K-means cluster con diferentes valores iniciales
- Se realizará T veces y se obtendrán T matrices
- Matrices binarias



| | | | |
|---|---|-----|---|
| 0 | 1 | ... | 0 |
| 1 | 0 | ... | 0 |
| ⋮ | ⋮ | ⋮ | ⋮ |
| 0 | 1 | ... | 0 |

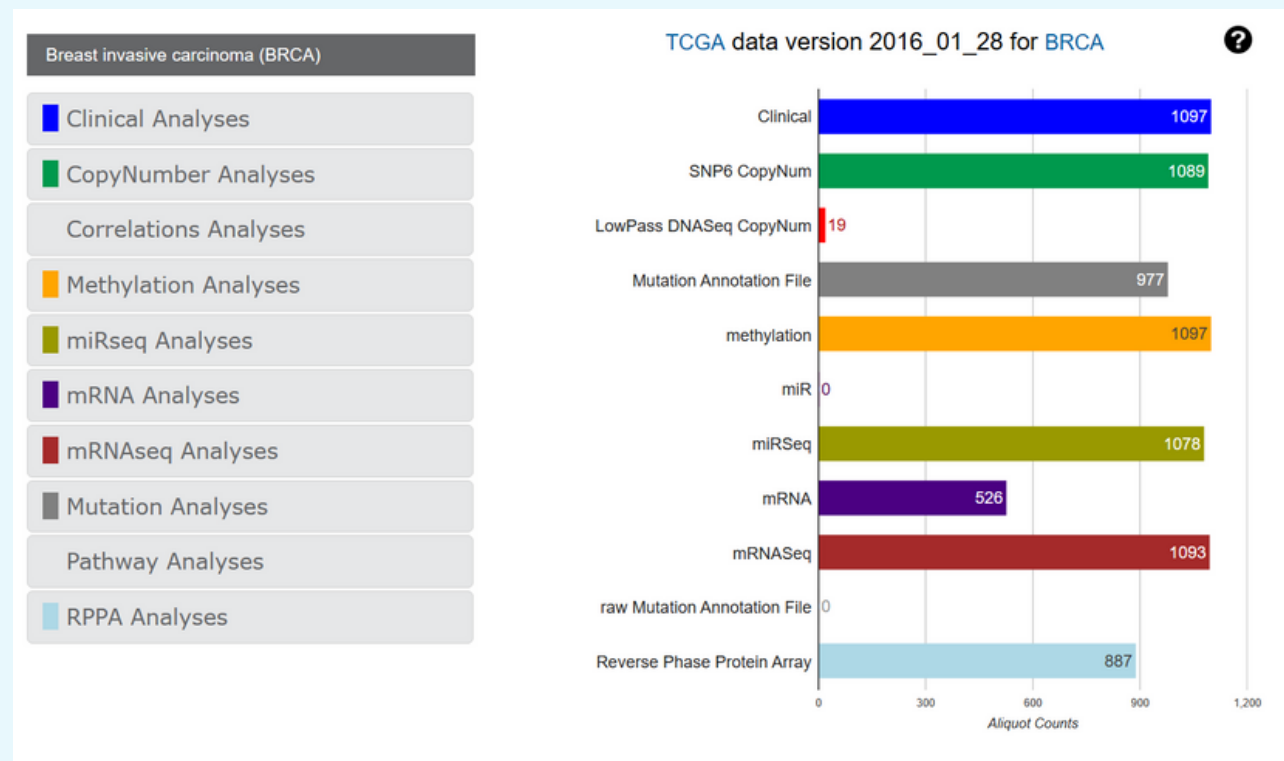
CONSENSO DE CLUSTERS

- Se concatenan las T particiones en una matriz
- Se le aplica el K-means cluster
- índice de Calinski-Harabaz
- Fusionamos los resultados en un consenso



Comparación,
conClustering y otros
algoritmos de agrupación

CONJUNTO DE DATOS

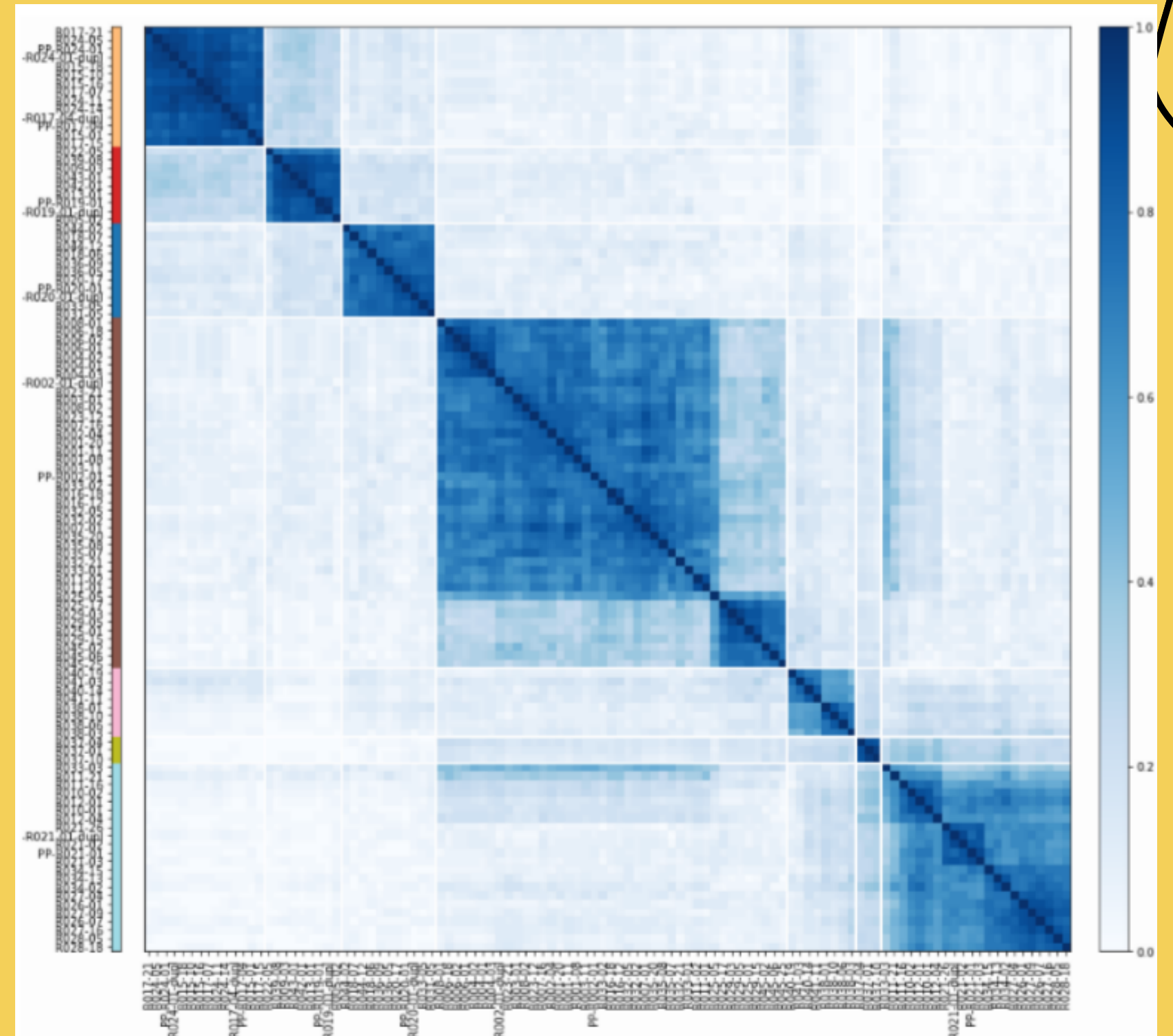


Los datos usados para probar el método fueron recuperados de la página FireBrowse y algunas secuencias de NLM. Este repositorio tiene en su conjunto los 1098 casos del TCGA del cancer de mama

RESULTADOS

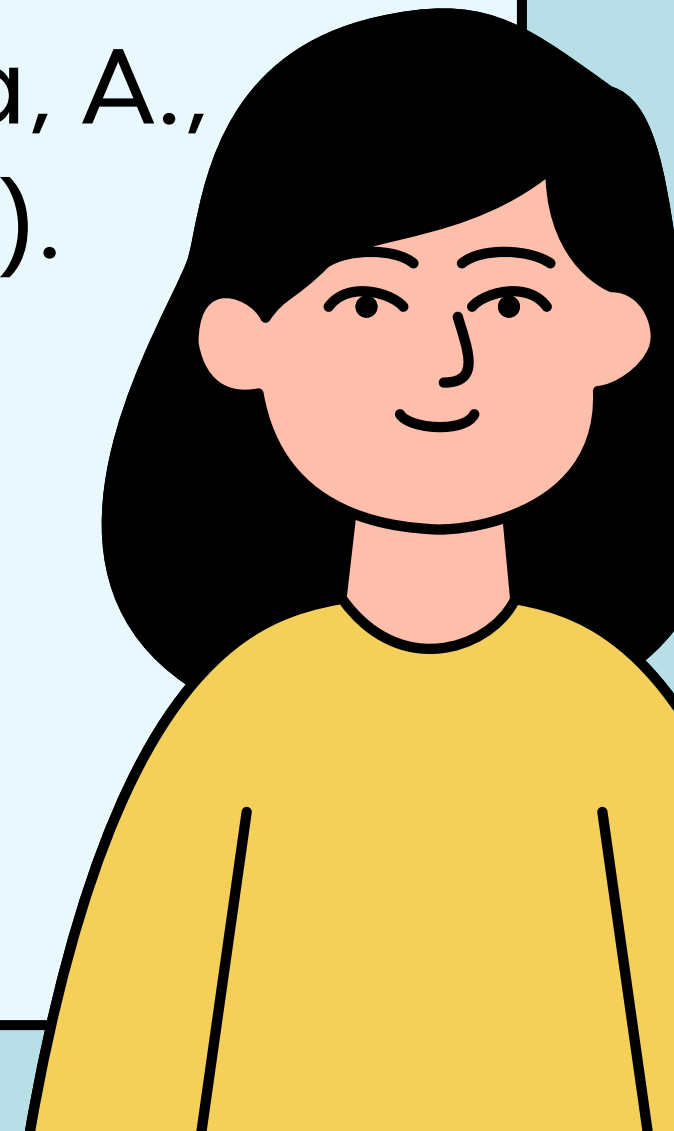
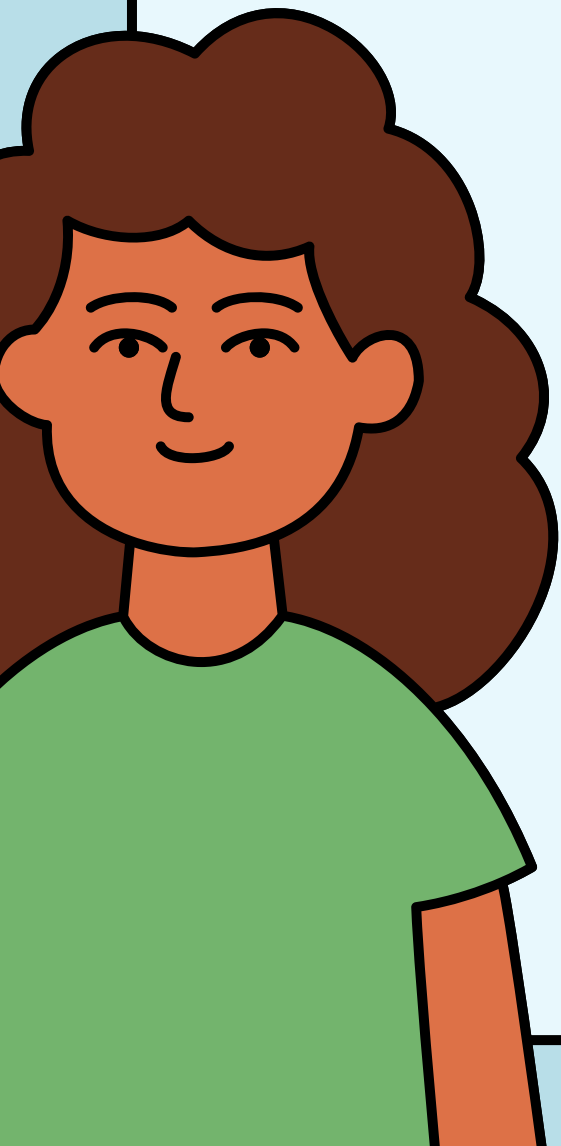
Consenso K-means

Agrupaciones de subtipos



Bibliografía

- Arzanova, E., & Mayrovitz, H. N. (2022). The Epidemiology of Breast Cancer. 1–20.
<https://doi.org/10.36255/exon-publications-breast-cancer-epidemiology>
- Sylwia Łukasiewicz, Marcin Czezelewski, Forma, A., Baj, J., Sitarz, R., & Andrzej Stanisławek. (2021). Breast Cancer—Epidemiology, Risk Factors, Classification,
- Broad Institute. (2019) TCGA data version 2016_01_28 for BRCA [FireBrowse](#)





GRACIAS POR SU ATENCIÓN