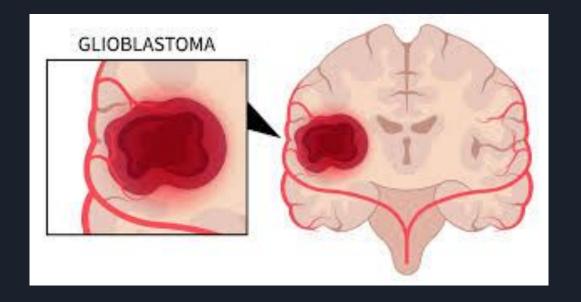
"Identificación de módulos y patrones funcionales mediante mutaciones en glioblastomas"

#### Elaborado por:

- Mier Fenogilo Sofía
- Pérez Romero Natalia Abigail
- Reyes Tapia Valeria
- Torres Aduna Sebastian Ulises

## Introducción

- ¿Qué es el cáncer?
- ¿Qué son los tumores?
- Glioblastomas
- Cáncer de piel



#### Introducción

Proyectos que trabajan con tumores han generado un gran volumen de datos sobre aberraciones genómicas, epigenómicas y de expresión génica.

Estos transforma nuestro concepto sobre la biología del cáncer, así como la posible revelación de nuevos biomarcadores, objetivos farmacológicos y la forma en la que posiblemente cambiará el desarrollo de nuevas terapias contra el cáncer.

## Pregunta de investigación

- "Discovering functional modules by identifying recurrent and mutually exclusive mutational patterns in tumors" publicado por Christopher Miller, Stephen Settle y Erik Sulman
- Técnicas que se han realizado en tumores así como las muestras obtenidas sólo con aberraciones genómicas, entre ellas, mutaciones puntuales y alteraciones del número de copias,
- Análisis anteriores han descubierto en una colección de muestras la alteración de genes que comprenden un módulo funciona y se han percatado que se encuentra en el mismo tumor, una explicación a esto, es que existen relaciones funcionales entre los genes (principalmente para los patrones RME).

## Objetivos

- Recrear una red de módulos funcionales con un método ya existente que permite identificar patrones de mutación recurrentes y mutuamente excluyentes
- Filtrar los módulos funcionales que afectan más en el cáncer de piel (melanomas) de forma computacional
- Eventualmente, poder obtener resultados de forma rápida y sencilla

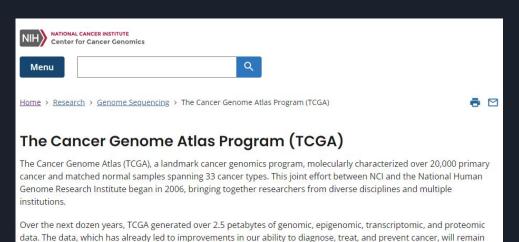
#### Métodos

- 1. Crear matriz de mutación
- 2. Construir una red de genes con el algoritmo Winnow
- 3. Identificar módulos candidatos
- 4. Evaluación de los módulos a través de una prueba de significancia algorítmica

#### Métodos.

publicly available for anyone in the research community to use.

En el artículo describe la metodología para obtener la matriz de mutación de la colección CGA de 145 muestras de glioblastoma.



"project id": "TCGA-GBM" "submitter id": "TCGA-02-0001", "submitter id": "TCGA-02-0001-01C", "sample type": "Primary Tumor", "specimen\_type": "Solid Tissue", "composition": "Not Reported", "updated\_datetime": "2023-11-08T11:20:37.055810-06:00", "analytes": [ "updated\_datetime": "2022-07-21T03:58:55.624657-05:00", "name": "Harvard Medical School",

#### Métodos

La línea de análisis GDC DNA-Seq identifica variantes somáticas en datos de secuenciación del exoma completo (WXS) y secuenciación del genoma completo (WGS). Las variantes somáticas se identifican comparando las frecuencias alélicas en alineaciones de muestras normales y tumorales, anotando cada mutación y agregando mutaciones de múltiples casos en un archivo de proyecto.

GDC_Aliquot Chromosome Start End Num	_Pro	bes Segment_Mean
228ce105-6d6e-401a-a504-d5c00c61be80	1	3301765 50148100 <b>25175</b> 0.0211
228ce105-6d6e-401a-a504-d5c00c61be80	1	50149844 50151878 3 -1.5846
228ce105-6d6e-401a-a504-d5c00c61be80	1	<b>50154650</b> 247650984 <b>104556</b> 0.0166
228ce105-6d6e-401a-a504-d5c00c61be80	2	480597 241537572 132182 0.0206
228ce105-6d6e-401a-a504-d5c00c61be80	3	2170634 197812401 107162 0.0173
228ce105-6d6e-401a-a504-d5c00c61be80	4	1059384 124538250 69507 0.016

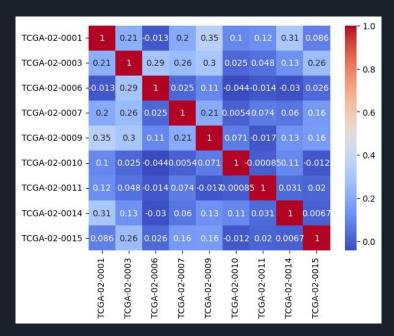
## Métodos

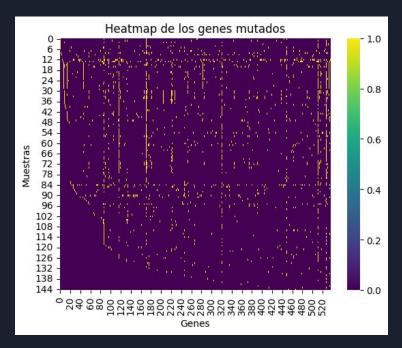
#### Matriz de mutación

TCGA-02-0001 TCGA-02-000					3	TC	GA-02	2-000	96	TC	GA-02	2-006	37	TC	TCGA-02-0009				TCGA-02-0010				GA-0:	2-001	TCGA-02-00					
ZNF708	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	8	0	0	0	0	0	ø	0	0	0	0	1	0	0
XPOT	0	0	Ø	0	0	0	0	0	0	0	0	0	Ø	Ø	0	0	0	0	0	0	0	0	Ø	0	0	0	0	0	0	0
RNF38	8	0	Ø	0	0	0	0	0	0	0	0	ø	Ø	9	0	0	ø	0	0	0	0	0	Ø	Θ	0	0	0	0	0	0
CCT2	8	0	ø	0	0	0	0	0	0	0	0	0	ø	0	ø	0	0	0	0	0	8	0	ø	0	0	0	0	ø	0	0
INSR	0	0	0	0	Ø	0	0	0	0	0	0	0	ø	0	0	0	9	0	0	0	0	0	ø	0	0	0	ø	0	0	0
ESR2	0	0	ø	0	0	0		0	0	0	0	0	Ø	0	0	0	8	0	0	0	0	0	ø	0	0	0	0	0	0	0
ATP1A2	0	0	0	0	0	0	8	0	0	0	0	0	Ø	0	Ø	0	8	0	0	0	0	0	0	0	0	0	8	ø	0	0
DGKD	0	0		0	0	0		0	0	0	0	0		0	0	0		0	0	0	0	0		0	0	0		0	0	0
chr6:16	8107	192-	1682	7676	93	0	1		1	0			0	1	0	0	0			0	1		0	0	0	0	0			0
GYPA		0			0	0		0	0	0		0			0	0		0	0	0		0			0	0		0	0	0
MN1 0	0		1	0	0		0	1	1		0		0	0			0	0	0	0	0		0	0			0	0	0	
RADIL		0		0	0	0		0	0	0		0		0	0	0		0	0	0	0	0			0	0		0	0	0
TRIM24		0		0	0	0		0	0	0	0	0			0	0		0	0	0		0			0	0		0	0	0
GYPB		0		0	0	0		0	0	0		0			0	0		0	0	0		0			0	0		0	0	0
PHIP		0		1	0			0	0	0		0			0	0		0	0	0		0		0	0	0		0	0	0
chr6:16	4016	427-	1646	5012	28		1		1	0	0		0	1	0	1	0			0	1		0	0	0	0	0			0
USF1		0		0	0	0		1	0	0		0			0	0		0	0	0		0		0	0	0		0	0	0

#### Resultados

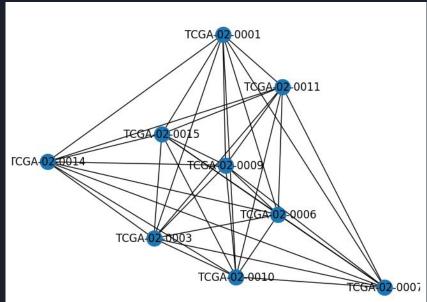
Calcular la puntuación de exclusividad entre cada par de genes, definida como el número de muestras en las que exactamente uno del par está mutado dividido por el número de muestras en las que al menos uno del par está mutado.

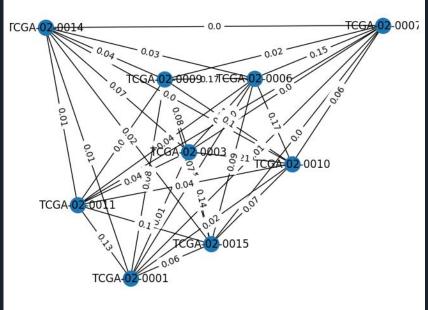




#### Resultados

Este análisis puede ayudar a identificar genes que están en rutas mutualmente exclusivas, lo que puede sugerir que las mutaciones en estos genes están proporcionando ventajas similares a las células tumorales y que una mutación en uno puede hacer innecesaria una mutación en el otro.





### Algoritmo de Winnow

Utilizamos el algoritmo Winnow para construir una red de genes a partir de los datos de mutaciones obtenidos de melanomas, con el objetivo de identificar patrones de aberraciones recurrentes y mutuamente excluyentes (RME).

Primero, creamos una matriz de mutación donde las filas representan muestras de pacientes y las columnas representan genes, con valores de 1 y 0 indicando la presencia o ausencia de mutaciones. De igual forma, utilizamos cada gen como clasificador y el resto de la matriz como datos de entrenamiento.

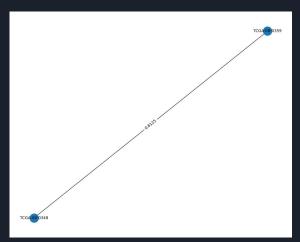
Ahora lo explicaremos lo que se realizó por ejecución:

Primera ejecución: Invertimos los bits de la matriz para predecir las aberraciones basadas en genes no mutados, ajustando los pesos de las aristas en consecuencia.

Segunda ejecución: Invertimos los bits del clasificador para predecir lo contrario, permitiéndonos así ajustar los pesos para considerar tanto las co-mutaciones como la exclusividad mutua. Estos pesos se ajustan multiplicativamente para resaltar las conexiones significativas y reducir o eliminar las conexiones o relaciones que podrían ser menos relevantes o menos significativas.

Siguiendo con el algoritmo construimos la red de genes, con nodos que representan genes y aristas basadas en los pesos calculados, eliminando aquellas con puntuaciones bajas para mantener solo las conexiones más significativas.

Finalmente, analizamos la red para identificar módulos de exclusividad mutua, permitiéndonos descubrir rutas críticas en el desarrollo del melanoma y potenciales dianas terapéuticas.



#### Identificación de Módulos Funcionales

Al realizar nuestro algoritmo de Winnow, encontramos la relación de los pares de genes con mejor puntaje de exclusividad.

Éste, nos servirá como punto de partida para el algoritmo de significancia.

```
sebastian@sebastian-BOM-WXX9:~/GC-2024-2/Provecto-Genomica-ComputacionalS pv
thon3 winnow algorithm.py
'TCGA-02-0016', 'TCGA-06-0156') 0.6744186046511628
                 'TCGA-06-0219') 0.5510204081632653
                 'TCGA-08-0348') 0.5625
                 'TCGA-06-0138') 0.55
                 'TCGA-06-0156')
                                 0.5609756097560976
                 'TCGA-06-0646') 0.6764705882352942
                 'TCGA-06-0648') 0.555555555555556
('TCGA-06-0138'
                 'TCGA-02-0048') 0.55
                 'TCGA-06-0156') 0.5454545454545454
                 'TCGA-06-0646') 0.6052631578947368
                 'TCGA-08-0358') 0.5217391304347826
                 'TCGA-06-0187') 0.52
                 'TCGA-08-0358') 0.6666666666666666
                 'TCGA-08-0375') 0.5185185185185185
                 'TCGA-02-0016') 0.6744186046511628
                 'TCGA-06-0138') 0.5454545454545454
                 'TCGA-06-0219') 0.574468085106383
                 'TCGA-06-0241') 0.55
                 'TCGA-06-0646') 0.5365853658536586
 'TCGA-06-0185'
                 'TCGA-08-0244') 0.5185185185185185
                 'TCGA-06-0154') 0.52
                 'TCGA-08-0353') 0.5416666666666666
                 'TCGA-08-0358') 0.5652173913043478
 'TCGA-06-0187'. 'TCGA-08-0375') 0.5416666666666666
                 'TCGA-06-0219') 0.5454545454545454
                 'TCGA-02-0016') 0.5510204081632653
('TCGA-06-0219', 'TCGA-06-0156') 0.574468085106383
```

# Evaluar los módulos por medio de una prueba de significancia algorítmica

- Puntaje de Exclusividad 0.5 y 0.6
- Pares de Genes con mejor relación de probabilidad.
- Fórmula Aplicada a la Teoría de Bayes.
- Algoritmo Nulo.

#### Paso 1. Codifica la matriz de aberración.

Se toma una variable d y se inicializa en 0. Recorremos la matriz fila por fila, en orden de izquierda a derecha, incrementando d de la siguiente manera en cada celda (i,j): Si x i,j = 1 entonces:

$$d' \leftarrow d' + (-\log(p_{NULL}(1)) + \log(p_{RME}(1))$$

en otro caso,  $d' \leftarrow d' + (-\log(p_{NULL}(0)) + \log(1 - p_{RME}(1))$ 

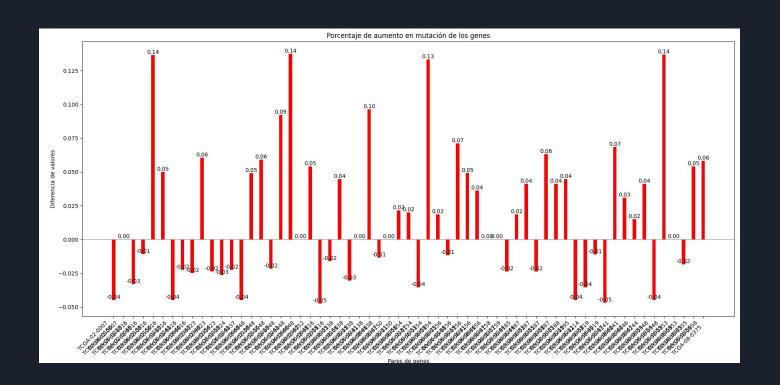
donde log denota logaritmo binario.

Paso 2 . Calcular la significancia.
 d = d' - m log(n) - k log\*(m) - m log\*(k)

$$\rho_{RME}(x_{i,j} = 1) \approx \rho(x_{i,j} = 1 | a_{i,j}, b_{i,j}) = \frac{\rho(x_{i,j} = 1 | a_{i,j}) \cdot \rho(x_{i,j} = 1 | b_{i,j})}{\rho(x_{i,j} = 1) \cdot \left(\frac{\rho(x_{i,j} = 0 | a_{i,j}) \cdot \rho(x_{i,j} = 0 | b_{i,j})}{\rho(x_{i,j} = 0)} + \frac{\rho(x_{i,j} = 1 | a_{i,j}) \cdot \rho(x_{i,j} = 1 | b_{i,j})}{\rho(x_{i,j} = 1)}\right)}$$

Figura 2. Modelo probabilístico para el cálculo de la siguiente distribución de aberración.

# Significancia Algorítmica



# Conclusiones

#### Con este análisis logramos:

- Descubrir patrones de exclusividad que pueden indicar interacciones genéticas específicas en el cáncer.
- Comparar la efectividad de este método con técnicas previas.
- Comparar los genes que posiblemente muten y las condiciones, así como analizar el cambio que se generan en éstas.
- El algoritmo previo permite analizar diversas formas de cáncer siempre y cuando se presenten en matrices de mutaciones proporcionadas por instituciones de confianza.

