

## Participantes:

1. José Luis Ayala (UCM)
2. Sandra Alonso Paz (UCM)

## Título:

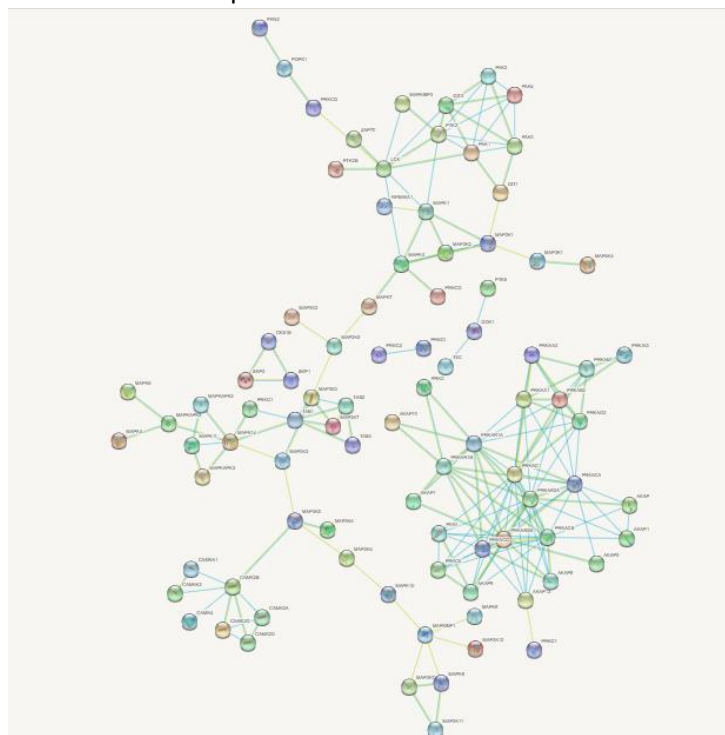
Tercera reunión individual.

## Vía:

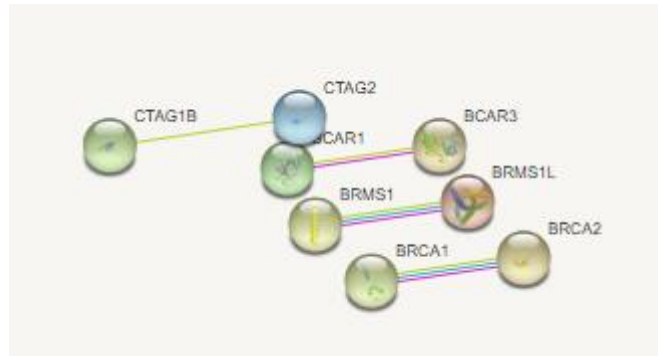
Online. Google Meet.

## Trabajo realizado desde la última reunión:

1. Revisar poster:
  - a. Quitar texto (no hacer frases completas si no más puntos).
  - b. Hay una imagen que he sacado de internet, no sé si afecta en algo el copyright. La puedo diseñar yo a las malas.
  - c. Queda rellenar autores: rellena él.
2. Búsqueda de genes:
  - a. Buscar entre los 43.893 genes de la base de datos. Filtrado:
    - i. Buscamos en Ensembl: de 47167, 33231 (70.5%) tienen información. (13935 (29.5%) son NaN) **(Busqueda más exhaustiva)**
    - ii. De 33231, 836 non-protein coding (2%). Lo que hace un total de 32.405.
    - iii. Si quitamos los genes repetidos por tener más de una entrada en ensembl, nos quedan 27.396. **(filtred\_genes.xlsx)**
    - iv. Como en String solo caben 2000 genes para hacer los grafos, guardamos solo los genes que sí codifican proteínas. Estos son 6281. **(only\_protein\_genes.xlsx)**
    - v. Si además quitamos las proteínas (no tiene propiedades descritas) (311), nos quedan 5965 genes. **(genes without uncharacterized proteins.xlsx)**
      1. Como una de nuestras principales dianas es la kinasa. Seleccionamos todas las proteínas relacionadas con la misma. Esto da 234 genes. Con un grado de confianza alto salen dos componentes.



2. Como también está relacionado con el cáncer: Esto da 58 genes. No da resultado.



- b. Comparar la cohorte que usaron en el paper de comienzo de las prácticas con los 144 genes encontrados en DisGeNet para Mrcc. Coger 13 genes y hacer modelo de clasificación para comparar.
    - i. Coincidencias: 13.
    - ii. Son firmas de genes ya creadas.
  - c. Mostrar con String y Huri las redes resultantes.
    - i. En Huri no carga la red (demasiado grade, 183 nodos).
    - ii. Con String solo deja hacer redes de 2000 proteínas y tenemos 5965. No sé por qué descartar\filtrar.
  - d. Buscar asociaciones efectivas como Cancer pathway.
3. Generación de grafos:
    - a. Usar Pytorch.
      - i. Ya he empezado a montar el gnn.
    - b. Probar con GraphGym.

## Trabajo para la siguiente reunión:

1. Poster:
  - a. Esperar confirmación de los oncólogos.
  - b. Preguntar quiénes serán autores finales.
  - c. Buscar autoría de la imagen del poster.
2. Probar la GNN con los 234 genes de kinasa.
3. Como segunda opción probar con los 5965 genes codificantes de proteínas. En este caso probar con Cytoscape.
 

([https://www.researchgate.net/post/How can I add more than 2000 data in STRINGS website](https://www.researchgate.net/post/How_can_I_add_more_than_2000_data_in_STRINGS_website))
4. Buscar otras asociaciones como cancer pathway.
5. Crear nueva rama en repo GitHub.