Reunión 4: 23/09/2022

## Participantes:

- 1. José Luis Ayala (UCM)
- 2. Sandra Alonso Paz (UCM)
- 3. Carlos Moral Rubio (UCM) -
- 4. María (UCM) Samur (Ictus)
- 5. Fernando (UCM) Afasia y neurodegeneración
- 6. Laura Hernández Lorenzo (UCM) Alzheimer por biomarcadores
- 7. Josué Ortiz (UPM)
- 8. Javier Gálvez Goicuria (UCM) Estudio del sueño
- 9. Paula Domínguez (UCM)

## Título:

Segunda reunión grupal.

#### Vía:

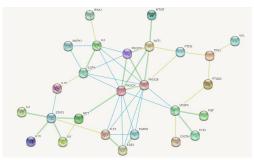
Presencial, Facultad de Informática.

## Trabajo hecho hasta la reunión.

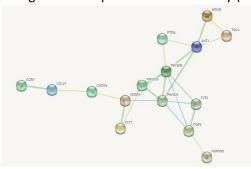
- 1. Ya he resuelto la mayoría de comentarios del paper que hizo Josué, pero tengo algunas preguntas (apuntadas en overleaf):
  - Duda mayúsculas y minúsculas.
  - Cambio de nombre de variables (Responde CB)
- 2. Poster propuesto:
  - Acabado. Revisar el contenido porque no he citado todo lo que hicimos porque es casi imposible que entre.
- 3. Búsqueda de genes entre los que podamos establecer una relación: (113 HURI) buscar en paper anterior!
  - Intenté filtrar los genes del Excel, pero no tuve muchos resultados y fue muy costoso. La mayoría eran desconocidos: solo se sabia su posición en el genoma y si codificaban proteínas. Pero que no codifiquen proteínas no significa que no estén interviniendo en otras funciones. BIOGRID
  - DisGeNet:
    - Busqué en DisGeNet la colección de genes asociados a mRCC. Aparecieron 144 genes que también estaban contenidos en nuestra base de datos (Excel).
      - https://www.disgenet.org/browser/0/1/0/C4721698/0/25/144/\_a/\_b./-score/
        - o Problemas:
          - 1. A diferencia del estudio de Laura, al filtrar por curated no sale ninguno. Los míos son todos "BEFREE".
          - 2. Ninguno de los genes relacionados con la respuesta inmune extraídos de las feature selection aparecen en esta base de datos (IL25, TRIM43B y SMG8).
    - Después, de estos 144 busqué de que formas los podíamos relacionar (además de por su conexión física). He encontrado varias formas, pero creo que la mejor podría ser por:
      - o protein domain (tyrosine kinase) (4-4) menores de 5



Kegg pathway: cancer pathway (25)

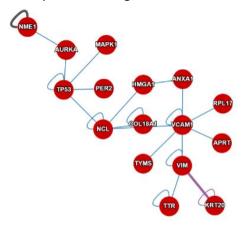


o Biological proces: regulatrion of protein kinase activity (15)



### • Huri: (la más fiable)

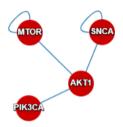
- He buscado los 144 genes y solo aparecen 113.
- La componente más grande es de 13



- o Proteine domain "tyrosine kinase": (14/14): no hay components conexas
- o Con el kegg pathway "cancer pathway": (28/32): componente más grande es de 3



 Proceso biológico "regulation of proteine kinase activity" (22/24): componente más grande es de 4



4. Lectura del paper de Laura. Tengo preguntas:

Reunión 4: 23/09/2022

- Predicción de genotipo a fenotipo que utiliza GNN.
- ¿Qué son los fenotipos asociados a la enfermedad del Alzheimer? ¿Son los síntomas?
- ¿Qué son las variantes de sentido erróneo?
- ¿Por qué en DisGeNet filtras por curated?
- 5. José se reunía hoy con los oncólogos. ¿Alguna novedad? NUEVA BD
- 6. ¡Hoy es mi primer día de trabajo!
  - La fecha de inicio del contrato es el 23 de septiembre.

## Trabajo hecho hasta la reunión.

#### 1. Pasos hechos:

- 1) Buscar genes asociados a mrcc: 144
- 2) Extraer el mapeo de las variantes genéticas (posición en el cromosoma) utilizando coordenadas genéticas GRCh37.
- 3) Usé String para ver qué relación se establecía entre ellos. https://string-db.org/cgi/network?taskid=bmyjhX810ox0&sessionId=bWKTItAQqhsx
  - Network type: physical network.
  - Meaning of network edges: evidence
  - Active interaction sources: textmining and databases
  - Interaction score: 0.4
  - Network display options: hide disconnected nodes and show input protein names.
- 2. Estudio de otras posibles relaciones entre genes: https://string-db.org/cgi/setevidence?taskid=bfls.BslfV803&sessionid=bWKTltAQqhsx&data\_channel=database
  - a. **Tissue expresión (tejido donde se expresa):** entre las primeras clasificaciones específicas encontramos:
    - Sistema urogenital (67).
    - Sistema reproductivo (62)
    - Órganos genitales femeninos (44).
  - b. **Biological process:** En este apartado no soy muy experta y no sé qué procesos biológicos incluyen este tipo de cáncer. Sin embargo, he encontrado procesos relacionados con la respuesta inmune y la regulación de la kinasa. Necesitaría ayuda para ver que procesos están involucrado y tiene sentido estudiar.
    - Inmune system process (68).
    - Inmune response (54).
    - Regulation of inmune system process (47).
    - Regulation of kinase activity (32).
    - Regulation of protein kinase activity (24).
  - c. Molecular function:
    - Binding (120).
    - Kinase activity (22).
  - d. Kegg pathways (interacciones moleculares dentro de las células) Impoortante
    - Pathways in cancer (32).
    - EGFR tyrosine kinase inhibitor resistance (17).
  - e. Reactome pathways
    - Inmune system (53).
    - Signaling by receptor tyrosine kinase (18).
  - f. Protein domain:
    - Tyrosine kinase (14).
- 3. Estudio GNN: PYTORCH
  - a. Aun no entiendo cuál es el input del grafo.
    - Matriz de adyacencia.
  - b. <a href="https://github.com/microsoft/gated-graph-neural-network-samples">https://github.com/microsoft/gated-graph-neural-network-samples</a>
  - c. <a href="https://ieeexplore.ieee.org/stamp/stamp.jsp?tp=&arnumber=4700287">https://ieeexplore.ieee.org/stamp/stamp.jsp?tp=&arnumber=4700287</a>

- https://github.com/sailab-code/GNN tf 2.x/tree/main/GNN
- https://github.com/tensorflow/gnn

# Trabajo para la siguiente reunión

No centarnos en los 144 genes porque sería replicar el trabajo del paper de Braun.

Hacer una búsqueda más exhaustiva de la descripción de los genes.

Revisar el poster para resumir el texto.

### HURI GENES (113)

EPHA2 ADIPOR2 AZIN2 VIM TIMP2 PKLR FLT4 SNCA FOXP3 UBASH3B PIK3CB RBM45 PRDM6 MS4A1 SLC12A2 SPATA2 NTN1 ADIPOR1 FGFR2 CYP3A4 TFE3 VCAM1 CXCR1 PITX1 EOMES APOB ABCB1 CTLA4 ALB FOLH1 PBRM1 MMP2 PTHLH IL15 **AURKA** CSF2 GALNT10 DYNLL1 DLC1 ICAM1 MAPK1 IFI27 JAG1 AXL STAT3 FLT1 TSC2 IL13 DCTN6 CXCL8 PIK3CG GJB1 MET KRT20 SERPINE1 PIK3CD PTEN PSMD9 ABO VEGFA TYMS IL4 ULK1 NCL ADIPOQ SETD2 TTR ABCG2 COL18A1 EIF4EBP1 PGF CD274 IFNA2 PDCD1 CD80 HLA-DRB1 PIK3CA CXCR4 ANXA6 FLT3 SMUG1 PELI1 PAX8 C5AR1 SERPINB2 KDR FGF13 PEG3 ACLY MTOR PER2 APRT HLA-DOA CRP EPHB2 IFNA13 NME1 ANXA1 TICAM2 HILPDA RPL17 DAB2IP

HLA-DQA1 CCL21 HMGA1 FGF2 ACVRL1 MFAP1 TP53 AKT1