



# Caracterització de les vies moleculars de la migranya i obtenció de productes gènics candidats

**Adrià Fernàndez Díaz**  
Bioinformàtica i Bioestadística  
Area 4 Aula 1

**Nom del consultor:** Daniel Aguilar Villalba  
**Nom de la professora responsable de l'assignatura:** David Merino Arranz

Data d'Entrega: Gener 2021



Esta obra está sujeta a una licencia de Reconocimiento-NoComercial-SinObraDerivada [3.0 España de Creative Commons](https://creativecommons.org/licenses/by-nc-nd/3.0/es/)

## FITXA DEL TREBALL FINAL

<b>Títol del treball:</b>	Caracterització de les vies moleculars de la migranya i obtenció de productes gènics candidats
<b>Nom de l'autor:</b>	Adrià Fernàndez Díaz
<b>Nom del consultor/a:</b>	Daniel Aguilar Villalba
<b>Nom del PRA:</b>	David Merino Arranz
<b>Data d'entrega (mm/aaaa):</b>	01/2021
<b>Titulació:</b>	Bioinformàtica i bioestadística
<b>Àrea del Trabajo Final:</b>	Àrea 4 Aula 1
<b>Idioma del treball:</b>	Català
<b>Paraules clau</b>	Migraine disorder, Interactome, Network Science.
<b>Resum del Treball (màxim 250 paraules):</b>	
<p>La migranya és una malaltia extesa entre la població amb uns símptomes que si be no són severs dificulten la vida de les persones que ho pateixen (nàusees, fotofòbia, fonofòbia i mal de cap). Actualment hi ha poc coneixement sobre els mecanismes moleculars involucrats i per tant els possibles tractaments són també reduïts. És per això que aquest treball proposa un estudi bioinformàtic per extreure informació rellevant per a la malaltia.</p> <p>El treball ha estat estructurat en dos blocs. Un primer bloc on s'ha executat un procés d'obtenció de dades per elaborar l'interactoma, obtenir les vies cel·lulars de l'<i>Homo sapiens</i>, i els gens que afecten la migranya que ja es coneixen experimentalment. El segon bloc consta de l'estudi estadístic d'aquestes dades obtingudes amb la finalitat d'obtenir un llistat amb mecanismes moleculars candidats a la migranya, i un altre llistat amb gens candidats.</p> <p>S'han obtingut 68 vies candidates i un conjunt de 30 gens candidats.</p> <p>Tot i això, alguns dels mecanismes cel·lulars i gens candidats obtinguts guarden relació amb el que ja se sap sobre la migranya, per tant podrien ser bones vies per futures investigacions.</p>	

**Abstract (in English, 250 words or less):**

Migraine is a widespread disease among the population with symptoms that, although not severe, make life difficult for people who suffer from it (nausea, photophobia, phonophobia and headaches). There is currently not enough knowledge about the molecular mechanisms involved and therefore the possible treatments are also reduced. That is why this research proposes a bioinformatics study to extract information relevant to the disease.

This study has been structured in two blocks. The first block is a process of obtaining data to elaborate the interactome. Also to obtain the cellular pathways of *Homo sapiens* and the genes that affect migraine that are already known experimentally. The second block consists of the statistical study of the data in order to obtain a list with molecular mechanisms candidate for migraine, and another list with candidate genes.

68 candidate pathways and a set of 30 candidate genes have been obtained.

However, some of the cellular mechanisms and candidate genes obtained are related to what is already known about migraine, so they could be good avenues for future research.

# Índex

1. Introducció.....	1
1.1 Context i justificació del treball.....	1
1.2 Objectius del treball.....	1
1.3 Enfocament i mètode seguit.....	2
1.4 Planificació del treball.....	2
1.5 Breu sumari dels productes obtinguts.....	3
1.6 Breu descripció dels altres capítols de la memòria.....	4
2. Resta de capítols.....	4
2.1 Materials i mètodes.....	4
2.1.1- Bloc 1. Obtenció i tractament de les dades:.....	4
2.1.2- Bloc 2. Càlculs estadístics sobre les dades:.....	9
2.2 Resultats:.....	12
2.2.1- Bloc 1. Obtenció i tractament de les dades:.....	12
2.2.2- Bloc 2. Càlculs estadístics sobre les dades:.....	14
2.3 Discussió:.....	17
2.3.1 Bloc 1. Obtenció i tractament de les dades:.....	17
2.3.2 Bloc 2. Càlculs estadístics sobre les dades:.....	18
3. Conclusions.....	20
4. Bibliografia.....	22
5. Annexos.....	24

# 1. Introducció

## 1.1 Context i justificació del treball

La migranya és una malaltia que afecta aproximadament al 10% de la població, amb uns símptomes com ara nàusees, fotofòbia, fonofòbia i mal de cap<sup>1</sup>. Per altre banda, els processos i mecanismes moleculars d'aquesta malaltia no són prou coneguts, i això dificulta la possibilitat de trobar teràpies exitoses i definitives, havent, actualment tractaments genèrics per cadascun dels tractaments, sent pal·liatius i no cures<sup>2</sup>. Els estudis d'aquest tipus poden suposar un cost per als investigadors massa elevat, i és aquí on intervé la bioinformàtica i el «network science» que ens permeten acotar el marc d'estudi. Alguns exemples d'aquesta aplicació seria l'estudi del trastorn depressiu major<sup>3</sup> o de les malalties respiratòries cròniques<sup>4</sup>.

En els darrers anys la bioinformàtica ha contribuït en el coneixement de malalties a nivell molecular permetent tractar i analitzar el gran volum de dades obtingudes (Big Data)<sup>5</sup>. El «network science», concretament, es basa en l'estudi de les connexions (interaccions moleculars) entre nodes (gens) generant, en el cas de la investigació genètica, l'interactoma.

Aquest treball, i totes les possibles variants que es poden obtenir per a diferents malalties, resulten de gran utilitat de cara a la recerca de nous tractaments.

Aquest treball es basa en l'estudi bioinformàtic dels gens associats a la migranya. Partint com a base de l'interactoma humà (les relacions establertes entre els gens), i mitjançant un estudi estadístic dels gens coneguts de la migranya, busquem com a objectiu final caracteritzar millor aquesta malaltia a nivell molecular i obtenir possibles productes gènics candidats a estar relacionats amb aquesta.

## 1.2 Objectius del treball

El treball està estructurat seguint dos objectius finals:

- 1- Obtenció d'informació sobre els mecanismes moleculars de la migranya.
- 2- Obtenció del llistat productes gènics candidats.

### **1.3 Enfocament i mètode seguit**

La metodologia seguida en aquest projecte està molt estudiada, i podem veure-la en multitud d'estudis que presenten grans similituds amb aquest treball<sup>3,4,5</sup>.

- Es partirà d'un procés de «data mining» per a obtenir, en primer lloc, l'interactoma humà. L'interactoma és el conjunt d'interaccions moleculars d'un conjunt que bé podria ser un organisme complet, o una cèl·lula concreta. En el nostre cas s'ha fet servir un interactoma de les relacions entre gens de tot el genoma humà. Això ens permetrà observar per una part les interaccions produïdes entre els gens de la migranya, i per altre part les interaccions entre els gens de les diferents vies mencionades en el següent apartat i la possible relació d'aquests amb els gens de la migranya. Aquesta informació s'obtindrà a partir de les diferents bases de dades públiques.
- Paral·lelament s'obtindrà també informació sobre les vies del metabolisme humà. La qual podrem comparar amb l'interactoma obtingut coneixent les connexions intra vies i permetent fer els estudis posteriors de la relació de la migranya amb aquestes.
- Seguirem amb un procés de «data mining» que ens permetrà obtenir la informació disponible a la literatura sobre els gens implicats en la migranya.
- La informació obtinguda a partir del «data mining» s'haurà d'harmonitzar per tal d'assegurar la homogeneïtat entre gens.
- A continuació s'executarà el procés estadístic del treball, començant amb un estudi de la connectivitat entre els gens de la migranya.
- Seguidament s'executarà un procés estadístic entre els gens de la migranya i les vies obtingudes, amb l'objectiu de trobar possibles mecanismes candidats.
- Finalment farem un anàlisi amb l'interactoma i la informació obtinguda sobre la migranya per obtenir el llistat d'altres productes gènics.

### **1.4 Planificació del treball**

Les tasques necessàries per aquest treball es poden desglossar en dos blocs:

Bloc 1, Data mining:

- Obtenció de dades per elaborar l'interactoma humà.
- Obtenció de les vies principals del metabolisme humà.
- Obtenció d'informació sobre els gens implicats en la migranya.

## Bloc 2, Estadística:

- Estudi estadístic sobre el grau d'associació dels gens de la migranya a les diferents vies cel·lulars. Obtenint així una taula amb els mecanismes cel·lulars potencialment relacionats amb la migranya classificats en funció del seu p-valor.
- Estudi sobre el grau d'associació dels gens de la migranya a la resta de gens de l'interactoma. Obtenint així una taula amb els gens potencialment relacionats amb la migranya classificats en funció del nombre de interaccions amb els gens de la migranya.

Aquestes tasques han estat planificades com es mostra a la figura 1, seguint les dates marcades per la UOC.

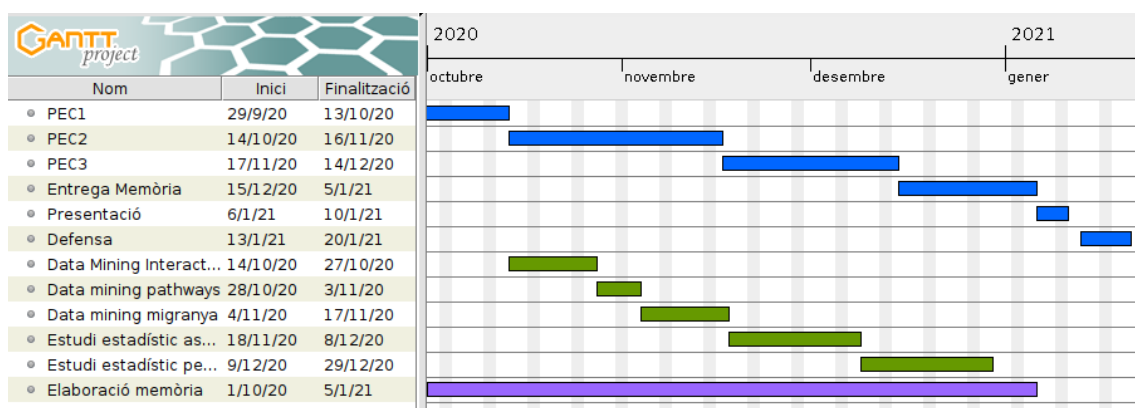


Figura 1: Diagrama de GANTT representant el pla de treball d'aquest TFM.

### 1.5 Breu sumari dels productes obtinguts

Els productes finals obtinguts formen dos blocs:

- Un primer on s'obtenen totes les dades que després es faran servir (interactoma, vies moleculars i gens de la migranya).
- Un segon on s'obtenen unes taules per mesurar estadísticament l'associació de la migranya amb els diferents mecanismes cel·lulars i una taula amb la classificació dels possibles gens candidats associats amb la migranya.



## **1.6 Breu descripció dels altres capítols de la memòria**

En el capítol 2 d'aquesta memòria estan continguts els apartats típics d'un article científic:

- Materials i mètodes, on s'exposa com s'ha executat el treball, detallant-ne els passos i procediments.
- Resultats, on s'exposen els productes obtinguts amb els procediments de materials i mètodes.
- Discussió, on s'analitzen els resultats obtinguts.

El capítol 3 correspon a les conclusions que s'extreuen del treball, tant a nivell de contingut com personal.

## **2. Resta de capítols**

### **2.1 Materials i mètodes**

Per a l'elaboració d'aquest estudi s'ha fet servir com a principal eina bioinformàtica el llenguatge de programació R<sup>6</sup> tant per a la part del tractament de les dades com per als estudis estadístics.

El treball s'ha englobat en dos blocs. El primer es basa en l'obtenció de dades i el tractament d'aquestes triant les dades d'interès i adequant el format per tal d'homogeneïtzar-ho entre les diferents fonts de partida. El segon bloc engloba el conjunt de càlculs estadístics.

#### **2.1.1- Bloc 1. Obtenció i tractament de les dades:**

A l'hora de triar les bases de dades s'han fet servir aquelles d'accés públic amb alta rellevància en el món de la investigació (estan citades a nombrosos articles). En totes elles, per referenciar els gens s'ha decidit fer servir el format de «gene symbol», també conegut com «nom vulgar» del gen. Per altra banda, es faran servir només les dades obtingudes experimentalment, ja que si féssim servir dades computacionals acumularíem l'esbiaix de les dades i el dels nostres càlculs.

## A) Interactoma:

Base de Dades de APID<sup>7</sup> (Agile Protein Interactomes DataServer):

A l'hora de triar quina base de dades fer servir s'ha tingut en compte la traçabilitat de la seva informació, ja que és important assegurar que la procedència de les dades no sigui computacional, aquest fet fa que s'hagin descartat altres bases de dades (com per exemple Biogrid<sup>8</sup>) i finalment s'hagi triat APID.

APID és una base de dades que proporciona una col·lecció completa d'interactomes de proteïnes per a més de 400 organismes basada en la integració d'interaccions físiques proteïna-proteïna (IPP) demostrades experimentalment. La construcció dels interactomes es fa amb un enfocament metodològic per informar dels nivells de qualitat i la cobertura dels proteomes per a cada organisme inclòs. D'aquesta manera, APID proporciona interactomes d'organismes específics. En conjunt, APID conté un total de 90.379 proteïnes diferents i 678.441 interaccions.

De l'apartat de interactomes s'ha seleccionat l'organisme *Homo sapiens* i la base de dades amb interaccions amb 2 evidències experimentals per interacció<sup>9</sup>.

- Tot i que la descarrega ja ha estat feta cercant l'interactoma humà, algunes de les entrades són de interaccions entre gens humans i gens d'altres organismes, i és per això que s'han filtrat per mantenir només les interaccions entre gens humans. En aquest procés s'ha passat de 171.448 interaccions a 137.880 (disminució del ~20%).
- De les dades s'han fet servir concretament les que feien referència als gens en format «gene symbol». Però en algunes interaccions aquestes caselles estaven buides. Per a solucionar això s'han aïllat aquells casos que no contenien referència en format «gene symbol». Mitjançant el codi d'Uniprot i gràcies a l'eina de conversió entre noms<sup>10</sup>, de la mateixa web de Uniprot, s'han obtingut els noms en format «gene symbol». En aquest procés s'han perdut 465 interaccions (<0,5%)
- A continuació s'han eliminat els possibles duplicats que podria haver i les auto interaccions. En aquest procés s'ha passat de 137.415 interaccions a 133.905 (disminució del ~3%).

- Finalment s'han generat dos arxius, un que conté tot l'interactoma processat, i l'altre que conté només els noms de tots els gens del nostre interactoma [Annexos 1 i 2].
- El codi d'execució d'aquest apartat es pot consultar a l'annex 3.

## **B) Vies cel·lulars:**

Base de Dades de Reactome<sup>11</sup>:

Reactome és una base de dades de vies i reaccions biològiques seleccionades manualment i revisades mitjançant «peer review». Reactome defineix una «reacció» com qualsevol esdeveniment de la biologia que canvia l'estat d'una molècula biològica. La unió, l'activació, la translocació, la degradació i els esdeveniments bioquímics clàssics que impliquen un catalitzador són reaccions.

D'aquesta base de dades s'ha fet servir l'arxiu de «physical entity identifier mapping files»<sup>12</sup>, concretament el que inclou totes les vies, en format Uniprot.

- Aquesta descarrega conté gens de diferents organismes, així que s'ha filtrat per mantenir només aquells provinents del *Homo sapiens*. Després d'aquest procés obtenim 2.425 vies.
- Els mètodes d'obtenció de dades d'aquest arxiu es poden dividir en la referència IEA (Inferred from Electronic Annotation) que no assegura la procedència de les dades, i la referència TAS (Traceable Author Statement) que són dades bibliogràfiques demostrables. Per evitar dades calculades computacionalment s'han mantingut només les TAS. En aquest procés s'ha passat de 2.425 vies a 2.336 (disminució del ~4%).
- Les dades en format «gene symbol» conté especificacions sobre el gen que no són d'interès per aquest treball (modificacions post-traduccionals, cadenes concretes, ...). Partint de l'identificador d'Uniprot s'ha fet servir l'eina de conversió entre noms<sup>10</sup> per generar el nom del gen aïllat.
- La conversió de noms pot haver generat nombrosos duplicats, així que s'ha efectuat un filtratge de duplicats. En aquest procés no s'han perdut vies, però sí que s'han eliminat redundàncies.

- Seguidament s'han eliminat els gens que no formaven part del nostre interactoma. En aquest procés s'han eliminat algunes vies passant de 2.336 a 2.253(disminució del ~4%)
- A continuació s'han eliminat aquelles vies que contenien menys de 10 gens associats, seguint l'estudi de Vijay et al.<sup>13</sup>. En aquest procés s'ha passat de 2.253 vies a 1.216(disminució del ~46%).
- Finalment s'han generat dos arxius, per una part el que conté les vies amb els gens associats, i per altre un arxiu amb les equivalències de conversió entre via-ID i el nom complet de la via [Annexos 4 i 5].
- El codi d'execució d'aquest apartat es pot consultar a l'annex 6.

Observant les vies englobades a Reactome s'ha detectat que mancaven processos moleculars concrets (conté els termes més genèrics) i estructures cel·lulars. Per això s'ha ampliat aquest apartat amb una segona base de dades.

Base de dades de Gene Ontology<sup>14,15</sup>:

Gene ontology (GO) és un consorci creat per desenvolupar un model computacional actual i complet de sistemes biològics, des del nivell molecular fins a vies més grans, sistemes cel·lulars i organismes.

La seva base de dades és una representació computacional del coneixement científic actual sobre les funcions dels gens de molts organismes diferents.

GO descriu les seves dades respecte tres ontologies, conjunt de coneixements dins d'un domini determinat:

a) Funció molecular: Processos d'escala molecular realitzades per productes gènics.

b) Component cel·lular: Estructures cel·lulars o el seu entorn extracel·lular on es produeix una determinada funció.

c) Procés biològic: Processos formats per múltiples activitats moleculars i per tant més globals.

D'aquesta base de dades s'ha fet servir l'arxiu «goa\_human.gaf»<sup>16</sup>, corresponent a la espècie de Homo sapiens i al conjunt de dades de proteïnes.

- D'entre els diferents vies d'obtenció de les dades, s'han triat només aquells resultats basats en mètodes experimentals, que segons la definició dels Evidence Codes per part de GO, són: EXP, IDA, IPI, IMP, IGI, IEP, HTP, HDA, HMP, HGI, HEP. [Annex 7]. En aquest procés s'ha passat de 18.430 vies a 11.999 (disminució del ~35%).
- A l'hora de seleccionar les dades d'interès, en aquest cas s'ha fet servir el format de «gene symbol», el codi GO per a les vies, i el tipus d'ontologia al qual pertany cada via, que ens permet estudiar-les per separat.
- S'han eliminat tant els duplicats que poden sorgir dels diferents mètodes d'obtenció de les dades, i també s'han descartat aquells gens que no formen part del nostre interactoma. En aquest procés s'ha eliminat gens redundants però s'han mantingut totes les vies.
- Seguidament s'han eliminat els gens que no formaven part del nostre interactoma. En aquest procés s'han eliminat algunes vies passant de 11.999 vies a 11.389 (disminució del ~5%).
- A continuació s'han eliminat aquelles vies que contenien menys de 10 gens associats, seguint l'estudi de Vijay et al.<sup>13</sup>. En aquest procés s'ha passat de 11.389 vies a 1,313 (disminució del ~88%).
- Finalment s'han generat tres arxius, corresponents a les vies separats en funció de la seva ontologia [Annexos 8,9 i 10].
- Paral·lelament s'ha generat un arxiu amb els equivalències entre el codi GO i el nom complet de la via [Annex 11]
- El codi d'execució d'aquest apartat es pot consultar a l'annex 12.

### **C) Migranya:**

Base de Dades de CTD<sup>17</sup> (Comparative Toxicogenomics Database):

CTD és una base de dades que proporciona informació seleccionada manualment sobre les interaccions substància química-gen/proteïna, les relacions substància química-malalties i les relacions entre gens i malalties. Aquestes dades s'integren amb dades sobre funcions i vies per ajudar a desenvolupar hipòtesis sobre els mecanismes subjacents a les malalties influenciades pel medi ambient.

A partir del cercador de gens<sup>18</sup>, s'han buscat tots aquells gens relacionats amb la migranya amb la paraula clau "migraine disorders", que formin part de l'*Homo sapiens*, i assegurant-nos que són gens demostrats experimentalment i no pas predits cercant només els "curated".

- A partir del «gene symbol», s'han eliminat aquells que no formaven part del nostre interactoma. En aquest procés s'ha passat de 19 gens a 16 (disminució del ~16%).

- S'ha generat l'arxiu corresponent per a la seva posterior consulta. [Annex 13]

- El codi d'execució d'aquest apartat es pot consultar a l'annex 14.

### **2.1.2- Bloc 2. Càlculs estadístics sobre les dades:**

Els càlculs estadístics s'han elaborat en tres línies, una primera part on s'ha estudiat la connectivitat dels gens de la migranya entre si. Una segona part amb l'objectiu de relacionar la migranya amb el conjunt de vies, per obtenir un llistat de candidates relacionats amb al migranya. Per últim s'ha elaborat un llistat de nous gens candidats per a la migranya.

#### **A) Connectivitat dels gens de la migranya:**

Primer de tot s'ha elaborat una taula amb els gens de la migranya que interaccionaven amb cadascun dels gens de la migranya. A continuació s'ha calculat el «Jaccard Overlap» entre els diferents gens de la migranya obtenint així un valor mitjà. Aquest càlcul permet mesurar el grau de similitud entre dos conjunts de gens calculant la fracció de gens veïns comuns entre aquests. En aquest cas s'ha calculat de dues maneres diferents. La primera pren com a divisor el total de gens veïns entre A i B, valorant els gens en comú entre el total de gens veïns. La segona, en canvi, pren com a divisor el mínim de gens veïns entre A i B i amb això valora els gens veïns en comú d'aquell gen que té menys gens veïns. Aquest segon càlcul permet donar millor valors en aquells casos en que hi ha molta diferencia entre el nombre de veïns però la majoria dels veïns d'un dels gens estan en el conjunt de l'altre.

Aquests càlculs estan descrits per les següents fórmules:

- $O_t = \frac{|A \cap B|}{|A \cup B|}$
- $O_m = \frac{|A \cap B|}{|\min(A, B)|}$

$O_t$ = Jaccard Overlap dividit pel total de gens veïns.  $O_m$ = Jaccard Overlap dividit pel mínim de gens veïns.  
 $A$ =Gens veïns de A.  $B$ = Gens veïns de B.

A la taula 1 s'exposa un exemple simplificat d'aquest càlcul:

Gen	Gens veïns	Jaccard Overlap total ( $O_t$ )	Jaccard Overlap mínims ( $O_m$ )
A	P,Q,R,S,T	1/3	2/3
B	R,T,U		

Taula 1: Exemple simplificat del càlcul del Jaccard Overlap ( $O_t$  i  $O_m$ ).

Per calcular la significació estadística dels dos valors de  $O$  ( $O_t$  i  $O_m$ ) per a la migranya, s'han fet 1000 iteracions calculant el  $O$  de 16 gens triats al atzar d'entre tot el conjunt de gens de l'interactoma, obtenint una distribució de valors per atzar. Finalment s'ha calculat el P-valor de  $O$  de la migranya per a les dues formules diferents [Annex 15 i 16].

El codi d'execució d'aquest apartat es pot consultar a l'annex 17.

## B) Vies candidates:

Aquest apartat s'ha produït tot per duplicat, per una part les vies obtingudes de reactome, i per altre part els obtinguts de Gene Ontology (que a la vegada han estat dividits en funció de la ontologia a la que pertanyen).

- Primer de tots s'han separat les diferents vies en funció de si contenien gens de la migranya o no. A continuació s'han executat tests de «Fisher»<sup>19</sup> per al conjunt de vies que contenien la migranya. Obtenint així, la significància estadística de la presència de la migranya en aquestes vies.
- Com s'han elaborat una quantitat elevada de tests s'ha executat una correcció de «false discovery rate», concretament la de Benjamini & Hochberg<sup>20</sup> per corregir el p-value en funció del nombre de resultats.

- Per obtenir una taula amb tot el conjunt de vies s'han afegit les que no tenien gens de la migranya, assignant valors de p-value de 1.
- En el cas de les vies provinents de Gene Ontology en aquest punt s'han combinat les tres ontologies.
- Paral·lelament s'ha fet una alternativa estadística fent servir l'eina externa BinoX<sup>21</sup>. Un mètode que permet contrastar la interacció (en anglès «crosstalk») entre dos conjunts de gens (els gens de la migranya i cadascuna de les vies). Aquesta interacció entre dos conjunts és descrita com el grau de connexió entre ells, és a dir, el nombre d'associacions entre gens. BinoX s'encarrega de fer un anàlisi estadístic de la significància d'aquest «crosstalk». Mitjançant el principi del mostreig d'importància de Monte Carlo<sup>22</sup> es generen un conjunt de xarxes (interactomes) aleatoris a partir de l'interactoma proporcionat. Amb aquests paràmetres aleatoritzats s'estima la probabilitat de cadascuna de les observacions.
- En aquest punt s'han afegit a la taula aquelles vies que no tenen cap veí comú amb la migranya i per als quals Binox no ha executat tests assignant-los un p-value de 1.
- Finalment s'han afegit identificadors de la base de dades d'origen i s'han combinat els resultats de Reactome i Gene Ontology en un mateix fitxer, tant per als resultats dels tests de «Fisher» com dels resultats de Binox. [Annexos 18 i 19]
- El codi d'execució d'aquest apartat es pot consultar a l'annex 20 i 21.

### **C) Gens Candidats:**

A partir de l'interactoma, s'ha elaborat un llistat amb el nombre de interaccions que té cada gen (que no forma part de la migranya) amb gens de la migranya. [Annex 22]

El codi d'execució d'aquest apartat es pot consultar a l'annex 17.



## **2.2 Resultats:**

### **2.2.1- Bloc 1. Obtenció i tractament de les dades:**

#### **A) Interactoma:**

Una vegada finalitzat el tractament de la base de dades de APID, s'ha aconseguit un interactoma que consta de 15.071 nodes (corresponent al nombre de gens diferents) amb un total de 133.905 interaccions entre aquests.

#### **B) Vies:**

En el cas de les vies s'han obtingut dos conjunts de dades:

- 1- Per una banda el provinent de reactome, que després del processament descrit a materials i mètodes conté 1.216 vies diferents.
- 2- Per altre banda el provinent de Gene Ontology, que després de processat conté 1.313 vies diferents englobant les tres ontologies.

#### **C) Migranya:**

Després del tractament descrit a materials i mètodes s'han obtingut 16 gens relacionats amb la migranya (Taula 2, descripció dels gens obtinguda de genecards<sup>23</sup>).

<b>Gene symbol</b>	<b>Nom complet</b>	<b>Breu descripció del gen</b>
ATP1A2	ATPase Na <sup>+</sup> /K <sup>+</sup> Transporting Subunit Alpha 2	Proteïna de la família de les ATPases transportadores de cations de tipus P i a la subfamília de Na <sup>+</sup> / K <sup>+</sup> -ATPases.
CACNA1A	Calcium Voltage-Gated Channel Subunit Alpha1 A	Subunitat alfa-1A, que s'expressa predominantment en el teixit neuronal.
CALCA	Calcitonin Polypeptide Alpha	Related Hormones peptídiques calcitonina, pèptid relacionat amb el gen de la calcitonina i katalcicina mitjançant un splicing alternatiu específic del teixit de l'ARN de les transcripcions gèniques i la divisió de proteïnes precursors inactives.
EDNRA	Endothelin Receptor Type A	Receptor de l'endotelina-1, un pèptid que té un paper important en la vasoconstricció potent i de llarga durada.
ESR1	Estrogen Receptor 1	Receptor d'estrògens, un factor de transcripció activat per lligands compost per diversos dominis importants per a la unió hormonal, la unió de l'ADN i l'activació de la transcripció.
FHL5	Four And A Half Domains 5	LIM Proteïna expressada coordinadament amb l'activador del modulador d'elements que responen a l'AMPc (CREM). S'associen i confereixen una potent funció d'activació transcripcional
HTR2A	5-Hydroxytryptamine Receptor 2A	Receptor de la serotonina, un neurotransmissor amb molts rols
LRP1	LDL Receptor Protein 1	Related Membre de la família de proteïnes del receptor de lipoproteïnes de baixa densitat. Aquest receptor participa en diversos processos cel·lulars, inclosa la senyalització intracel·lular, l'homeòstasi lipídica i l'aclariment de les cèl·lules apoptòtiques.
MEF2D	Myocyte Enhancer Factor 2D	Membre de la família de factors de transcripció del factor potenciador de miocits específic 2 (MEF2). Els quals participen en el control de la diferenciació i el desenvolupament de les cèl·lules neuronals i musculars i estan regulats per les histones deacetilases de classe II.
PHACTR1	Phosphatase And Regulator 1	Actin Proteïna membre de la família de proteïnes reguladora de la fosfatasa i l'actina. Aquest membre de la família pot unir l'actina i regular la reorganització del citoesquelet d'actina.
PRDM16	PR/SET Domain 16	Factor de transcripció de dits de zinc i conté un domini PR terminal.
SCN1A	Sodium Voltage-Gated Channel Alpha Subunit 1	Subunitat alfa del canal de sodi, que té quatre dominis homòlegs, cadascun dels quals conté sis regions transmembrana
SUGCT	Succinyl-CoA:Glutarate-CoA Transferase	Proteïna similar als membres de la família de proteïnes CaiB / baiF CoA-transferasa
TAC1	Tachykinin Precursor 1	Aquest gen codifica quatre productes de la família de les hormones del pèptid taquiquinina, la substància P i la neurocinina A, així com els pèptids relacionats, el neuropèptid K i el neuropèptid gamma.
TGFBR2	Transforming Growth Factor Beta Receptor 2	Proteïna transmembranal que té un domini de proteïna quinasa, forma un complex heterodimèric amb el receptor TGF-beta tipus-1 i s'uneix al TGF-beta.
TNF	Tumor Necrosis Factor	Citocina proinflamàtòria multifuncional que pertany a la superfamília del factor de necrosi tumoral (TNF).

Taula 2: Llistat de gens pertanyents a la migranya, amb el seu nom complet i una breu descripció de la seva funcionalitat.

## 2.2.2- Bloc 2. Càlculs estadístics sobre les dades:

### A) Connectivitat dels gens de la migranya:

La connectivitat entre els gens de la migranya ha estat pràcticament nul·la, obtenint només una interacció entre LRP1 i CACNA1A (figura 2; versió amb més resolució a l'annex 23).

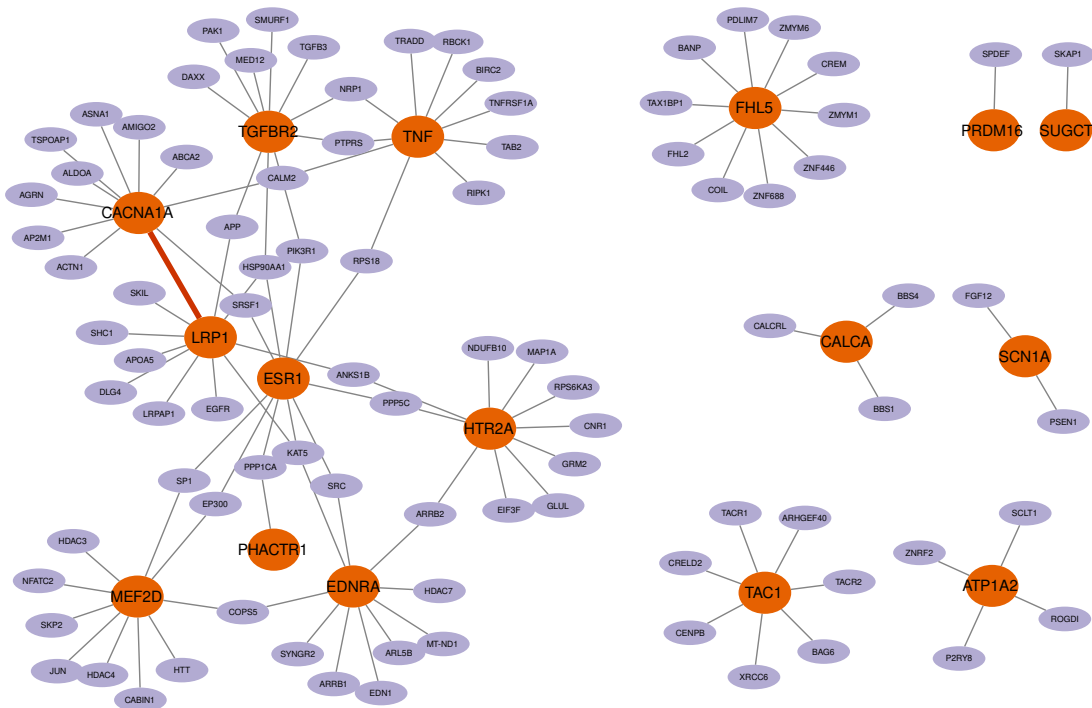


Figura 2: Representació via cytoscape de les interaccions dels gens de la migranya (taronja) i els 10 veïns (gris) de cada gen amb més relacions amb la migranya. Remarcada en taronja la interacció entre LRP1 i CACNA1A.

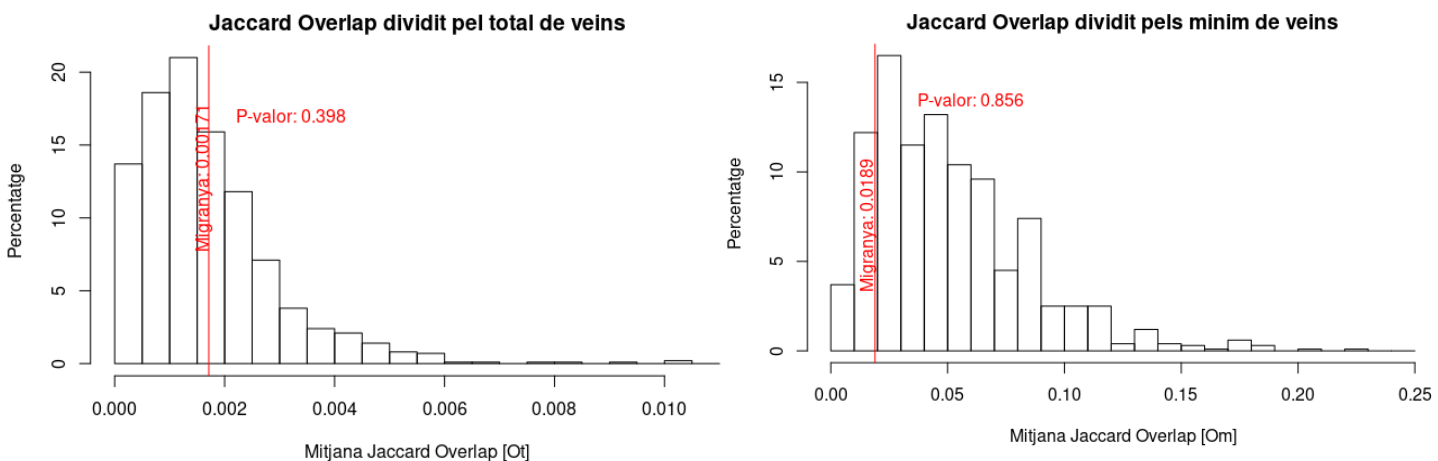


figura 3: Distribució aleatòria de  $O_i$  per a 1000 iteracions amb la superposició de  $O_i$  de la migranya 0,0171, amb un p-valor de 0,398.

figura 4: Distribució aleatòria de  $O_m$  per a 1000 iteracions amb la superposició de  $O_m$  de la migranya 0,0189, amb un p-valor de 0,856.

## B) Vies candidates:

La selecció de vies candidats s'ha elaborat a partir de dos mètodes diferents:

- Mitjançant els test de «Fisher» s'han obtingut 9 vies amb un p-valor ajustat inferior a 0,05 i per als quals, per tant, tenim significància estadística. [Annex 18]. A la taula 3 s'han disposat les vies així com la família en la qual s'engloben i el seu p-valor ajustat.

Nom de la via	Família en la que s'engloba la via	P-valor ajustat FDR
positive regulation of cytosolic calcium ion concentration	Calci	0,000595
Signal Transduction	Senyalització	0,004993
Signaling by GPCR	Senyalització	0,004993
GPCR downstream signalling	Senyalització	0,004993
GPCR ligand binding	Senyalització	0,004993
G alpha (q) signalling events	Senyalització	0,010416
Class A/1 (Rhodopsin-like receptors) response to drug	Comunicació cel·lular	0,021925
activating transcription factor binding1	Transcripció	0,034334
		0,047504

Taula 3. Llistat de vies amb significància estadística després del test de «Fisher» i la correcció FDR. Informació afegida sobre la família en la qual s'engloba cadascuna de les vies.

- Mitjançant Binox s'han obtingut un total de 59 casos amb un p-valor ajustat inferior a 0,05 i per als quals, per tant, tenim significància estadística. [Annex 19]. A la taula 4 s'han disposat les vies així com la família en la qual s'engloben i el seu p-valor ajustat.

Nom de la via	Família en la que s'engloba la via	P-valor ajustat FDR
Positive regulation of transcription by RNA polymerase II	Transcripció	0,000000
Transforming growth factor beta binding	Factor de creixement	0,000000
PPARA activates gene expression	Transcripció	0,000000
Regulation of lipid metabolism by PPARalpha	Lípids	0,000000
Transcription coactivator activity	Transcripció	0,000002
Transcriptional regulation of white adipocyte differentiation	Transcripció	0,000002
ESR-mediated signaling	Senyalització	0,000005
Signaling by Nuclear Receptors	Senyalització	0,000009
Estrogen receptor binding	Estrògens	0,000027
Metabolism of lipids	Lípids	0,000034
Thyroid hormone receptor binding	Tiroides	0,000050
Signal Transduction	Senyalització	0,000051
Activin binding	Activina	0,000064
Extracellular space	Comunicació cel·lular	0,000075
Nuclear receptor coactivator activity	Transcripció	0,000110
Developmental Biology	Genèric	0,000367
RNA binding	Genèric (RNA)	0,000410
Calcium ion binding	Calci	0,000691
Nucleus	Genèric (Nucli)	0,000699
Positive regulation of transcription, DNA-templated	Transcripció	0,001006
Viral mRNA Translation	Translació	0,001115
L13a-mediated translational silencing of Ceruloplasmin expression	Translació	0,001140
Formation of a pool of free 40S subunits	Translació	0,001261
Structural constituent of ribosome	Translació	0,001305
mediator complex	Transcripció	0,001549
Nonsense Mediated Decay (NMD) independent of the Exon Junction Complex (EJC)	NMD	0,001778
SRP-dependent cotranslational protein targeting to membrane	Translació	0,001997
collagen-containing extracellular matrix	Matriu extracel·lular	0,002108
GTP hydrolysis and joining of the 60S ribosomal subunit	Translació	0,002221
Eukaryotic Translation Initiation	Translació	0,002496
Cap-dependent Translation Initiation	Translació	0,002496
Translational initiation	Translació	0,002556
GPCR downstream signalling	Senyalització	0,002735
Heart development	Desenvolupament	0,002994
rRNA processing in the nucleus and cytosol	Translació	0,003093
Influenza Viral RNA Transcription and Replication	Transcripció	0,003572
Translation	Translació (Molt genèric)	0,003816
Focal adhesion	Comunicació cel·lular	0,003895
Nonsense-Mediated Decay (NMD)	NMD	0,004615
Nonsense Mediated Decay (NMD) enhanced by the Exon Junction Complex (EJC)	NMD	0,004615
Cytosolic large ribosomal subunit	Translació	0,004700
Signaling by GPCR	Senyalització	0,005059
Major pathway of rRNA processing in the nucleolus and cytosol	Translació	0,006046
rRNA processing	Translació	0,006731
Disease	Genèric	0,006821
SRP-dependent cotranslational protein targeting to membrane	translació	0,007441
Nuclear-transcribed mRNA catabolic process, nonsense-mediated decay	NMD	0,010295
heparin binding	Heparina	0,014873
Influenza Infection	Transcripció	0,015192
Chromatin binding	Genèric (Cromatina)	0,019759
RNA Polymerase II Transcription	Transcripció	0,022663
Heparan sulfate proteoglycan binding	Heparan	0,026380
Generic Transcription Pathway	Transcripció	0,027152
Extra-nuclear estrogen signaling	Senyalització	0,027951
Nucleoplasm	Genèric	0,032120
Estrogen-dependent gene expression	Estrògens	0,036184
extracellular region	Comunicació cel·lular	0,036311
Signaling by NTRKs	Senyalització	0,045814
Gene expression (Transcription)	Transcripció	0,047561

Taula 4. Llistat de vies amb significància estadística després del procés de BinoX i la correcció FDR. Informació afegida sobre la família en la qual s'engloba cadascuna de les vies.

### C) Gens Candidats:

S'ha obtingut un llistat amb el nombre de interaccions que té cada gen (que no forma part de la migranya) amb gens de la migranya [Annex 22]. S'observen 2 gens amb 3 interaccions (KAT5 i HSP0AA1) i 28 amb 2 interaccions (Figura 5), la resta tenen  $\leq 1$ .

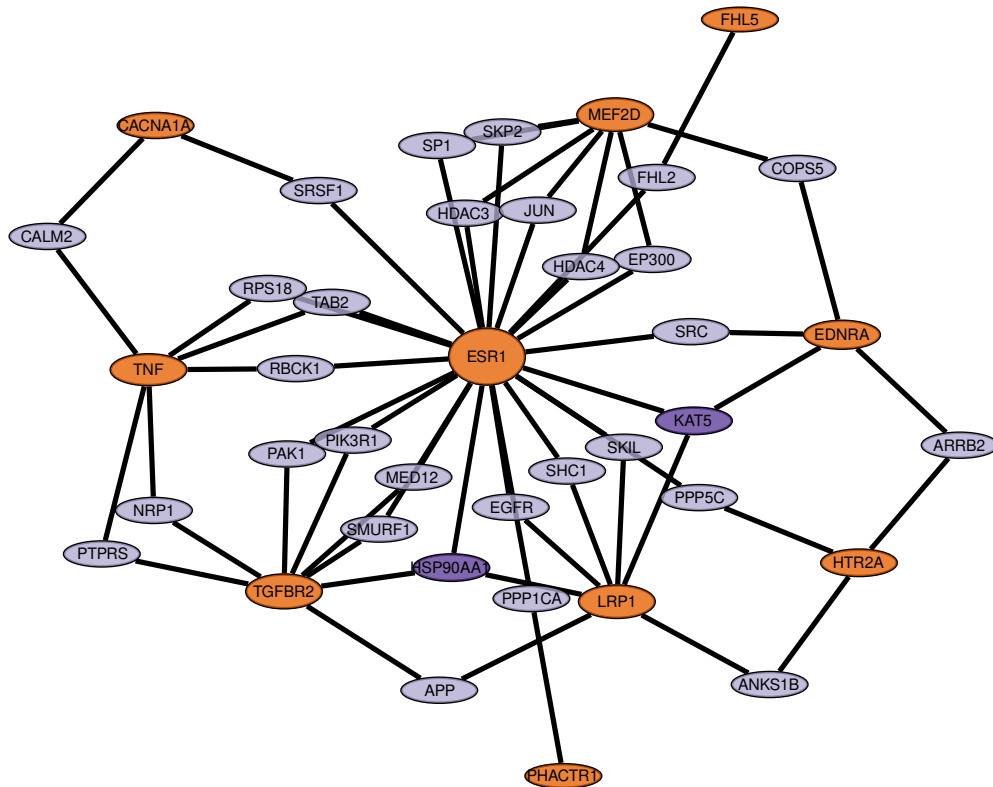


Figura 5: Representació via cytoscape dels gens de la migranya (taronja) i els 30 gens més relacionats amb la migranya amb 2 interaccions (blau clar) i amb 3 interaccions (lila fosc).

## 2.3 Discussió:

### 2.3.1 Bloc 1. Obtenció i tractament de les dades:

A l'hora de triar les dades s'han seguit dues vies principals de selecció:

- Per una banda s'ha decidit treballar amb interaccions entre proteïnes, en front de interaccions entre productes gènics en global. Això fa que es perdin tots els productes gènics no proteics (tals com el tRNA, el rRna, miRNA,...) a canvi d'assegurar que s'obtenen les variants proteiques dels gens.
- Per altra banda s'ha fet servir una selecció d'evidències sobre les dades que fos experimental. Això fa que els nostres resultats tinguin més robustesa i siguin més fiables, però a canvi d'una disminució del gruix de dades inicials i per tant s'està perdent informació que podria ser d'interès.

Aquests dos punts de triatge han fet que la selecció del nostre interactoma condicionés les bases de dades tractades tant de les vies com de la migranya:

- A l'hora de tractar les bases de dades de les vies s'ha tingut en compte l'interactoma i la referència de fer servir només les vies amb més de 10 gens associats<sup>13</sup>. S'han exclòs les vies amb menys gens ja que tenen poc sentit biològic. Aquest triatge ha fet que es perdessin aproximadament el 93% de les vies de Gene Ontology i això ha afectat al nostre contingut en resultats.
- Per altra banda amb els gens associats a la migranya partíem d'una quantitat reduïda, degut a que es tracta d'una malaltia poc estudiada. Això, sumat a les limitacions del interactoma, ha fet que passéssim de 19 gens a 16 gens, perdent el 15% de la informació inicial del nostre objectiu d'estudi.
- Per últim, Gene Ontology és una base de dades molt detallada a nivell de vies i sub-vies, això fa que sigui molt rica en contingut però a la vegada que sigui redundat.

### **2.3.2 Bloc 2. Càlculs estadístics sobre les dades:**

#### **A) Connectivitat dels gens de la migranya:**

Els resultats en els càlculs estadístics per valorar la connectivitat de la migranya han estat poc significatius. Per tant l'associació dels gens de la migranya no es estadísticament significativa així que aparentment no es pot afirmar una relació seguint l'interactoma. Comparant-ho amb estudis similars realitzats sobre altres malalties<sup>24,25,26</sup> s'observa que és un resultat atípic. Fet que aparentment demostra que els gens de la migranya no participen en els mateixos mecanismes moleculars.

Es plantegen tres hipòtesis per justificar aquesta falta de connectivitat:

- És habitual que en aquesta classe d'estudis es produeixi un esbiaix de coneixement. L'interactoma humà, tot i que en els darrers anys ha progressat molt, encara dista de ser complet al 100%<sup>27</sup>. D'igual manera la migranya no és de les malalties més estudiades<sup>28</sup>. Aquests dos conjunts de situacions fan que es parteixi de fonts poc completes i la manca d'informació en dificulti l'estudi i els resultats. Això pot provocar que l'estudi de la connectivitat estigui passant per alt algunes interaccions o gens que encara no estan documentats i per tant els resultats no siguin tan acurats.

- La migranya en sí es tracta d'una malaltia agrupada pels símptomes que genera<sup>29</sup>. Això podria fer que en el fons es tractessin de malalties relativament diferents i per tant amb vies diferenciades però que s'agrupen per simptomatologia.
- L'obtenció experimental de les dades de la migranya també podria ser un punt conflictiu. Els experiments fets servir per obtenir les evidències de relació entre els gens i una malaltia concreta sol produir-se a partir de la sang perifèrica, degut a al seva comoditat d'extracció i poca intrusivitat. La migranya, com d'altres malalties, està localitzada en un punt de difícil accés per prendre mostres, el cervell. Això pot propiciar el punt anterior de manera concreta ja que els les dades obtingudes poden ser incompletes.

### **B) Vies candidates a ser associades a la migranya:**

Estudis recents<sup>30</sup> han demostrat experimentalment l'associació de la migranya a un conjunt de vies moleculars. Concretament les vies relacionades amb el ritme circadià, els canals iònics, els mecanismes cel·lulars i sinàptics, els mecanismes vasculars, la homeòstasis dels ions metàl·lics i el metabolisme de la glucosa.

Dels resultats estadístics obtinguts [Annexos 18 i 19], i el triatge posterior de les taules 3 i 4 podem destacar-ne el següent:

- Les vies englobades a la família del calci, que a la vegada tenen relació amb els canals iònics i per tant en el funcionament de la sinapsi, ambdós prèviament ja relacionats a la migranya.
- Les vies englobades a la família de la comunicació cel·lular i senyalització que estan relacionats amb els mecanismes cel·lulars i sinàptics.
- Per altra banda caldria destacar l'elevada presència de vies relacionades amb la traducció i translació. Que si bé no s'ha trobat bibliografia demostrant la seva relació amb la migranya seria un bon punt de partida en futures recerques per ampliar el coneixement al voltant de la malaltia.

### **C) Els gens candidats associats a la migranya:**

Degut a la baixa connectivitat que s'ha observat a l'apartat 2.3.2.A, s'han obtingut pocs gens candidats, amb valors de nombre d'interaccions amb la migranya de  $\leq 3$ .



Analitzant<sup>23</sup> les vies associades amb els gens candidats obtinguts se'n destaca:

- L'elevada presència de gens relacionats amb al senyalització: HSP90AA1, RBCK1, SMURF1, TAB2, HDAC3, EGFR, JUN, CALM2, SRC, PIK3R1, HDAC4.
- Dos gens relacionats amb al comunicació cel·lular: APP, PTPRS.
- Un dels gens amb millor associació amb la migranya KAT5 (3 interaccions) forma part dels mecanismes moleculars relacionats amb el DNA i la cromatina. Uns processos que no han estat relacionats amb la migranya.

### 3. Conclusions

#### • Conclusions del treball:

Aquestes conclusions estan directament relacionades amb els objectius de l'apartat 1.2:

1-Referent als mecanismes moleculars aquest estudi ha trobat computacionalment un conjunt de vies que poden relacionar-se amb les ja conegudes per la literatura. I també s'ha trobat un conjunt de vies que aparentment no ho estan i podrien ser interessos per a futurs estudis.

2-Referent als gens candidats tot i que no presenten grans valors de connectivitat amb la migranya, s'ha trobat un conjunt de gens amb sentit biològic dintre la migranya, ja que intervenen en algunes vies relacionades amb la malaltia.

La falta de connectivitat s'ha de tenir en compte en la interpretació dels resultats referent als candidats trobats. Ja que redueix el valor estadístic d'aquests.

#### • Assoliment dels objectius:

L'apartat més informàtic del treball, de tractament de les dades i càlculs estadístics ha sigut enriquidor i complet. Per altra banda, l'apartat de discussió dels resultats ha resultat més breu del que podria haver estat, entrant dintre d'uns valors acceptables però sent possible una millor extensió.

- **Valoració de la planificació i metodologia:**

La metodologia seguida durant el treball ha sigut la correcte, ha permès arribar als objectius desitjats, tot i que la distribució de la feina no ha estat la idònia. S'ha notat falta de temps per elaborar millor l'apartat d'interpretació dels resultats.

- **Línies de recerca pendents:**

La línia de continuació de la recerca més interessant seria la d'estudiar molècules que tinguin com a diana algun dels gens predits, o alguna de les vies moleculars associades a la migranya. A partir d'aquestes molècules es podrien dissenyar tractaments per a la migranya.

## 4. Bibliografia

1. Dodick DW. A Phase-by-Phase Review of Migraine Pathophysiology. *Headache*. 2018;58 Suppl 1:4-16.
2. Ong JJY, De Felice M. Migraine Treatment: Current Acute Medications and Their Potential Mechanisms of Action. *Neurotherapeutics*. 2018 Apr;15(2):274-290.
3. Peilin Jia, Chung-Feng Kao, Po-Hsiu Kuo, Zhongming Zhao. (2011) . A comprehensive network and pathway analysis of candidate genes in major depressive disorder. *BMC Systems Biology* 2011, 5(Suppl 3):S12.
4. Enrico Maiorino et al. (2020). Discovering the genes mediating the interactions between chronic respiratory diseases in the human interactome. *NATURE COMMUNICATIONS* | 11:811 .
5. Sebastian Köhler, Sebastian Bauer, Denise Horn, and Peter N. Robinson. (2008). Walking the Interactome for Prioritization of Candidate Disease Genes. *The American Journal of Human Genetics* 82, 949–958.
6. R Core Team (2020). R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. URL <https://www.R-project.org/>.
7. Alonso-López D, Campos-Laborie FJ, Gutiérrez MA, Lambourne L, Calderwood MA, Vidal M, De Las Rivas J. APID database: redefining protein-protein interaction experimental evidences and binary interactomes. *Database* (Oxford). 2019; doi: 10.1093/database/baz005.
8. Oughtred R, Rust J, Chang C, Breitkreutz BJ, Stark C, Willems A, Boucher L, Leung G, Kolas N, Zhang F, Dolma S, Coulombe-Huntington J, Chatr-Aryamontri A, Dolinski K, Tyers M. The BioGRID database: A comprehensive biomedical resource of curated protein, genetic, and chemical interactions. *Protein Sci*. 2020 Oct 18.
9. <http://cicblade.dep.usal.es:8080/APID/init.action> [Consultat el 9.11.2020]
10. <https://www.uniprot.org/uploadlists/> [Consultat el 27.11.2020]
11. Jassal B, Matthews L, Viteri G, Gong C, Lorente P, Fabregat A, Sidiropoulos K, Cook J, Gillespie M, Haw R, Loney F, May B, Milacic M, Rothfels K, Sevilla C, Shamovsky V, Shorser S, Varusai T, Weiser J, Wu G, Stein L, Hermjakob H, D'Eustachio P. The reactome pathway knowledgebase. *Nucleic Acids Res*. 2020 Jan 8;48(D1):D498-D503. doi: 10.1093/nar/gkz1031. PubMed PMID: 31691815.
12. <https://reactome.org/download-data> [Consultat 23.10.2020]
13. Ramanan VK, Shen L, Moore JH, Saykin AJ. Pathway analysis of genomic data: concepts, methods, and prospects for future development. *Trends Genet*. 2012 Jul;28(7):323-32. doi: 10.1016/j.tig.2012.03.004. Epub 2012 Apr 3. PMID: 22480918; PMCID: PMC3378813.
14. Ashburner et al. Gene ontology: tool for the unification of biology. *Nat Genet*. May 2000;25(1):25-9.
15. The Gene Ontology resource: enriching a GOld mine. *Nucleic Acids Res*. Jan 2021;49(D1):D325-D334.
16. <http://current.geneontology.org/products/pages/downloads.html> [Consultat 26.11.2020]

17. Davis AP, Grondin CJ, Johnson RJ, Sciaky D, Wieggers J, Wieggers TC, Mattingly CJ. The Comparative Toxicogenomics Database: update 2021. *Nucleic Acids Res.* 2020 Oct 17
18. <http://ctdbase.org/> [Consultat 3.11.2020]
19. Fisher RA (1934) *Statistical methods for research workers*, 5th edn. Oliver & Boyd, Edinburgh
20. Benjamini Y, Hochberg Y (1995) Controlling the false discovery rate: a practical and powerful approach to multiple hypothesis testing. *J R Stat Soc B* 57:289–300
21. Ogris C, Guala D, Helleday T, Sonnhammer EL. A novel method for crosstalk analysis of biological networks: improving accuracy of pathway annotation. *Nucleic Acids Res.* 2017 Jan 25;45(2):e8. doi: 10.1093/nar/gkw849. Epub 2016 Sep 22. PMID: 27664219; PMCID: PMC5314790.
22. Siegmund D. Importance sampling in the Monte Carlo study of sequential tests. *Ann. Stat.* 1976;4:673–684.
23. <https://www.genecards.org/> [consultat 20.12.2020].
24. Menche J, Sharma A, Kitsak M, Ghiassian SD, Vidal M, Loscalzo J, Barabási AL. Disease networks. Uncovering disease-disease relationships through the incomplete interactome. *Science.* 2015 Feb 20;347(6224):1257601. doi: 10.1126/science.1257601. PMID: 25700523; PMCID: PMC4435741.
25. Goh KI, Cusick ME, Valle D, Childs B, Vidal M, Barabási AL. The human disease network. *Proc Natl Acad Sci U S A.* 2007 May 22;104(21):8685-90. doi: 10.1073/pnas.0701361104. Epub 2007 May 14. PMID: 17502601; PMCID: PMC1885563.
26. Xu J, Li Y. Discovering disease-genes by topological features in human protein-protein interaction network. *Bioinformatics.* 2006 Nov 15;22(22):2800-5. doi: 10.1093/bioinformatics/btl467. Epub 2006 Sep 5. PMID: 16954137.
27. Luck K, Sheynkman GM, Zhang I, Vidal M. Proteome-Scale Human Interactomics. *Trends Biochem Sci.* 2017 May;42(5):342-354. doi: 10.1016/j.tibs.2017.02.006. Epub 2017 Mar 8. PMID: 28284537; PMCID: PMC5409865.
28. Sutherland HG, Albury CL, Griffiths LR. Advances in genetics of migraine. *J Headache Pain.* 2019 Jun 21;20(1):72. doi: 10.1186/s10194-019-1017-9. PMID: 31226929; PMCID: PMC6734342.
29. Burstein R, Nosedá R, Borsook D. Migraine: multiple processes, complex pathophysiology. *J Neurosci.* 2015 Apr 29;35(17):6619-29. doi: 10.1523/JNEUROSCI.0373-15.2015. PMID: 25926442; PMCID: PMC4412887.
30. Rainero I, Roveta F, Vacca A, Noviello C, Rubino E. Migraine pathways and the identification of novel therapeutic targets. *Expert Opin Ther Targets.* 2020 Mar;24(3):245-253. doi: 10.1080/14728222.2020.1728255. Epub 2020 Feb 13. PMID: 32054351.

## 5. Annexos

- 1: «interactoma\_tractat.txt»
- 2: «nomsgens\_interactoma.txt»
- 3: «Obtencio\_interactoma\_APIID.RMD»
- 4: «pathways\_tractat\_reactome.txt»
- 5: «equivalencies\_pathways\_reactome.txt»
- 6: «Obtencio\_pathways\_reactome.Rmd»
- 7: «text suplementari.txt»
- 8: «pathways\_tractat\_GO\_P.txt»
- 9: «pathways\_tractat\_GO\_F.txt»
- 10: «pathways\_tractat\_GO\_C.txt»
- 11: «equivalencies\_pathways\_GO.txt»
- 12: «Obtencio\_pathways\_go.Rmd»
- 13: «gensmigranya\_tractat.txt»
- 14: «Obtencio\_migranya.Rmd»
- 15: «JaccardO\_migranya.csv»
- 16: «JaccardO\_migranyaV2.csv»
- 17: «Estadistics\_migranya.Rmd»
- 18: «resultats\_fisher\_all.csv»
- 19: «resultats\_binox\_all.csv»
- 20: «pathways\_migranya\_reactome.Rmd»
- 21: «pathways\_migranya\_GO.Rmd»
- 22: «llistat\_relacions\_migranya.csv»
- 23: «interaccions\_migranya\_triades.svg»