

PAC2 Desenvolupament del treball - Fase 1

Vasyl Druchkiv

Estudiant del Màster de Bioestadística i Bioinformàtica

15 d'Abril 2019

Índice

| | | |
|----------|---|-----------|
| 1 | Descripció de l'avenç del projecte | 2 |
| 2 | L'anàlisi comuna de GO, KEGG i Reactome | 5 |
| 2.1 | ORA | 5 |
| 2.1.1 | GO | 5 |
| 2.1.2 | KEGG | 6 |
| 2.1.3 | Reactome | 7 |
| 2.2 | GSEA | 8 |
| 2.2.1 | GO | 8 |
| 2.2.2 | KEGG | 9 |
| 2.2.3 | Reactome | 10 |
| 2.3 | Bar-Plots | 10 |
| 2.4 | Dot-Plots | 11 |
| 2.5 | Enrichment Plots | 12 |
| 2.6 | Category-Gene-Network Plot | 12 |
| 2.7 | GSEA Plot | 13 |
| 3 | L'anàlisi específic de GO, KEGG i Reactome | 14 |
| 3.1 | GO Plot | 14 |
| 3.2 | KEGG Pathway | 15 |
| 3.3 | Reactome Pathway | 15 |
| 4 | Validació dels resultats | 16 |
| 5 | Activitats no previstes | 23 |

1 Descripció de l'avenç del projecte

A data d'avui he desenvolupat l'aplicació d'anàlisi de les rutes. L'aplicació és completament funcional localment i ofereix l'anàlisi a partir de les bases de dades GO, KEGG i Reactome. A l'apartat **Input data** l'usuari primer ha d'indicar l'espècie per a totes tres bases de dades. Per les bases de dades GO i Reactome l'usuari pot elegir entre "human", "rat", "mouse", "celegans", "yeast", "zebrafish", "fly". Hi ha més espècies disponibles per a l'anàlisi KEGG, perquè la funció de `culsterProfiler enrichKEGG()` descarrega les últimes anotacions directament de la base de dades KEGG. Es poden trobar totes les espècies aquí. També l'usuari pot buscar l'espècie introduint els termes de cerca. Finalment l'usuari puja l'arxiu amb els gens i els LogRatios provinents de l'estudi de microarrays o NGS.

Figure 1: Pàgina d'entrada

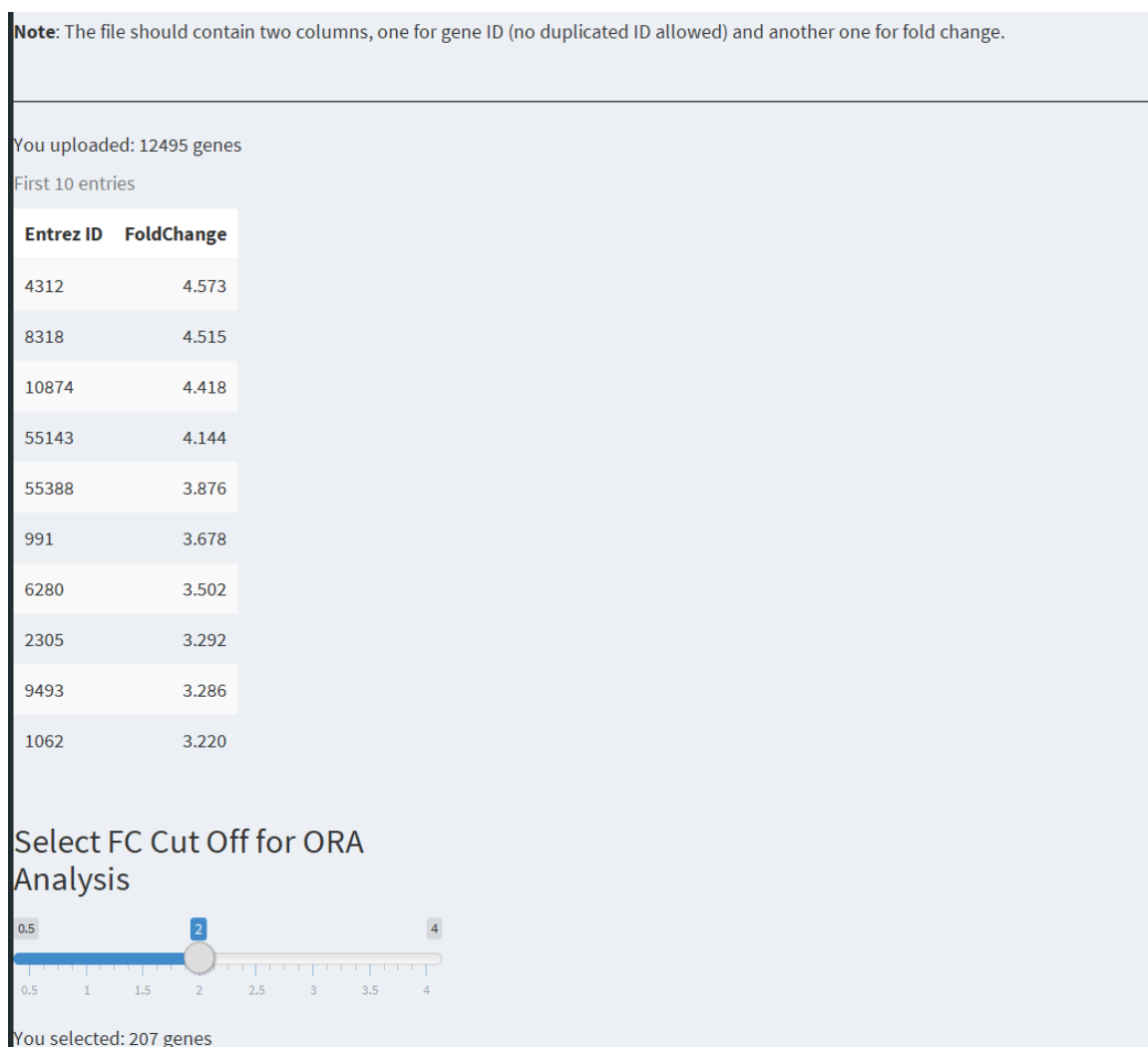
The screenshot shows the 'Pathway analysis' application interface. On the left is a dark sidebar with a menu: 'Data input' (with an upload icon), 'GO Analysis', 'KEGG Analysis', and 'Reactome Analysis'. The main content area has a light blue header with the title 'GO/REACTOME'. Below this, there is a 'Select Specie:' dropdown menu currently showing 'Homo Sapiens'. The next section is titled 'KEGG' and contains an 'Enter Search Term for Specie' input field with 'homo' entered, and a 'Select KEGG Specie' dropdown menu showing 'Homo sapiens'. Below that is a section titled 'File with all genes' with a 'Browse...' button and a 'No file selected' status. At the bottom, a note states: 'Note: The file should contain two columns, one for gene ID (no duplicated ID allowed) and another one for fold change.'

Un cop introduïdes les dades es mostra un petit resum del contingut i es dona la possibilitat d'elegir el *cut-off* de FoldChange per a l'anàlisi ORA. Per defecte s'agafa el valor de FoldChange=2. En funció del *cut-off* elegit es mostra el nombre dels gens en aquest grup (gens sobre o sotaexpressats).

L'aplicació està dividida doncs en 4 parts substancials:

1. Entrada de les dades;
2. Anàlisi GO;
3. Anàlisi KEGG;
4. Anàlisi Reactome.

Figure 2: El resum de les dades selecció del *cut off* per a l'anàlisi ORA



L'aplicació ofereix dos mètodes d'anàlisi: d'una banda es pot fer ORA (Over-Representation Analysis) i d'altra banda l'anàlisi GSEA (Gene Set Enrichment Analysis). Recordem que l'ORA consisteix a seleccionar els gens diferencialment expressats i basant-se en GO, KEGG o Reactome comprovar si una de les agrupacions de gens suggerides per aquestes bases de dades està sobre o sotraexpressada en els gens seleccionats. Per dur a terme l'ORA l'usuari té l'opció de definir un *cut-off* de Log-Ratio per formar el conjunt dels gens que s'hi utilitzarà (*gene set*). ORA és una bona eina per veure els efectes grans però els efectes petits se li escapen. Els efectes petits derivats dels gens individuals poden acumular-se en un efecte conjunt substancial el qual ORA no serà capaç de detectar. És aquí on GSEA mostra la seva utilitat.

Els apartats d'anàlisi (GO, KEGG i Reactome) ofereixen tan representacions comunes com representacions específiques.

Els anàlisis i representacions en comú són:

- Taula dels resultats ORA;
- Taula dels resultats GSEA;
- Gràfic de barres del resultat ORA;

- Gràfic de punts del resultat ORA;
- El mapa d'enriquement (Enrichment Map);
- La xarxa dels gens en categories (Category-gene-network);
- El gràfic de GSEA.

Les anàlisis específics són:

- GO → Gràfic GO
- KEGG → Rutes de la base de dades KEGG
- Reactome → Rutes de la base de dades Reactome

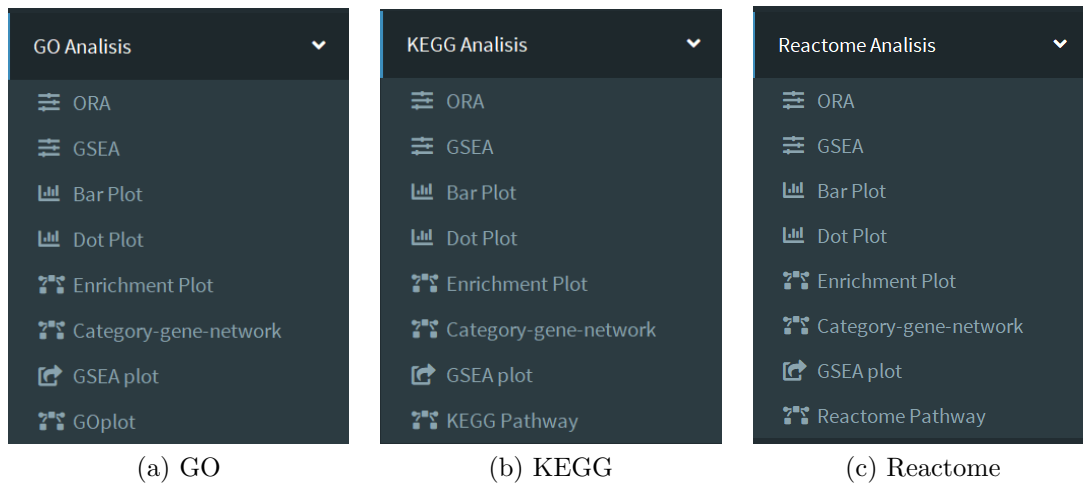


Figure 3: Els elements de les seccions d'anàlisi

| Base de dades | Mètode | Paquet Bioconductor | Funció | Observació |
|---------------|---------------------------|---------------------|-----------------|--|
| GO | ORA | clusterProfiler | enrichGO() | Només 7 espècies disponibles |
| GO | GSEA | clusterProfiler | gseGO() | Permutació de gens |
| GO | Bar-Plot | enrichplot | barplot() | Nececita l'objecte del class enrichResult |
| GO | Enrichment Map | enrichplot | emapplot() | Nececita l'objecte del class enrichResult |
| GO | Gene-Concept Network | enrichplot | cnetplot() | Nececita l'objecte del class enrichResult |
| GO | GO directed acyclic graph | enrichplot | goplot() | Nececita l'objecte del class enrichResult |
| KEGG | ORA | clusterProfiler | enrichKEGG() | Totes les espècies de KEGG |
| KEGG | GSEA | clusterProfiler | gseKEGG() | Permutació de gens |
| KEGG | Bar-Plot | enrichplot | barplot() | Nececita l'objecte del class enrichResult |
| KEGG | Enrichment Map | enrichplot | emapplot() | Nececita l'objecte del class enrichResult |
| KEGG | Gene-Concept Network | enrichplot | cnetplot() | Nececita l'objecte del class enrichResult |
| KEGG | Pathway | pathview | pathview() | Cal modificar la funció per guardar els gràfics en el directori temporal |
| Reactome | ORA | ReactomePA | enrichPathway() | Totes les espècies de KEGG |
| Reactome | GSEA | ReactomePA | gsePathway() | Permutació de gens |
| Reactome | Bar-Plot | enrichplot | barplot() | Nececita l'objecte del class enrichResult |
| Reactome | Enrichment Map | enrichplot | emapplot() | Nececita l'objecte del class enrichResult |
| Reactome | Gene-Concept Network | enrichplot | cnetplot() | Nececita l'objecte del class enrichResult |
| Reactome | Pathway | ReactomePA | viewPathway() | Molt lent. Per aquest motiu no he afegit botó de descarga. |

Table 1: Resum de les anàlisis disponibles i recursos de Bioconductor R

2 L'anàlisi comuna de GO, KEGG i Reactome

2.1 ORA

2.1.1 GO

Per realitzar l'anàlisi ORA per a termes GO s'utilitza la funció `enrichGO` del paquet `clusterProfiler`.

```
enrichGO(gene, OrgDb, keyType = "ENTREZID", ont = "MF", pvalueCutoff = 0.05,
pAdjustMethod = "BH", universe, qvalueCutoff = 0.2, minGSSize = 10, maxGSSize = 500,
readable = FALSE, pool = FALSE)
```

He implementat els valors per defecte amb la possibilitat per a usuari d'elegir entre:

- Ontologies GO
 - Molecular function, Biological proces, Cellular Components;
- Nivell de significació basant-se en els valors de P ajustats
 - 0.1, 0.05, 0.01, 0.001;
- Mètode d'ajustament
 - Holm; Hochberg; Hommel; Bonferroni; BH; BY; FDR; None.

L'execució de la funció és un procés temporalment costós. Per aquest motiu he afegit el botó d'acció, en lloc de deixar la funció reactiva. D'aquesta manera l'usuari ha de fer una decisió consient de repetir l'anàlisi amb altres valors.

Prement el botó apareix la taula i el botó nou mitjançant el qual l'usuari pot descarregar els resultats en format .csv. He formatjat la taula amb els paquets `knitr`, `kableExtra`, `formattable` i `dplyr`. Amb els dos últims he afegit les barres de color pel nombre dels gens diferencialment expressats del terme específic de GO i la gradació de color del verd fins al vermell pels valors dels més petits fins els més grans.

Els camps més interessants de la taula són:

- Description. El nom del terme GO;
- GeneRatio. El quocient:
$$\frac{\text{Nombre dels gens diferencialment expressats que pertanyen al conjunt de gens}}{\text{Nombre total dels gens diferencialment expressats}} = \frac{M}{N};$$
- BgRatio. El quocient:
$$\frac{\text{Nombre dels gens del conjunt d'interès en tota la mostra}}{\text{Nombre total dels gens en la mostra}} = \frac{k}{n};$$
- pvalue. Valor de p basat en la distribució hipergeomètrica:
$$p = 1 - \sum_{i=0}^{k-1} \frac{\binom{M}{i} \binom{N-M}{n-i}}{\binom{N}{n}}$$
- p.adjust. El valor de P ajustat.

The screenshot shows the 'Pathway analysis' interface. On the left is a dark sidebar with navigation options: 'Data input', 'GO Analysis' (selected), 'ORA', 'GSEA', 'Bar Plot', 'Dot Plot', 'Enrichment Plot', 'Category-gene-network', 'GSEA plot', 'GOplot', 'KEGG Analysis', and 'Reactome Analysis'. The main area is light blue and contains the following settings:

- Select ontology:** Radio buttons for CC, MF, and BP. BP is selected.
- Select adjustment method:** A dropdown menu showing 'BH'.
- Select P-Value threshold:** Radio buttons for 0.1, 0.05, 0.01, and 0.001. 0.1 is selected.
- A 'Calculate Result' button is at the bottom.

Figure 4: Especificació d'ORA dels termes GO

The screenshot shows the results table after clicking 'Calculate Result'. At the top, there is a 'Calculate Result' button and a 'Download Results as .csv' link. The table has the following columns: ID, Description, GeneRatio, BgRatio, pvalue, p.adjust, qvalue, Count, and geneID. Two rows are visible, both with green highlights in the p.adjust and Count columns.

| ID | Description | GeneRatio | BgRatio | pvalue | p.adjust | qvalue | Count | geneID |
|------------|--------------------------|-----------|-----------|--------|----------|--------|-------|---|
| GO:0140014 | mitotic nuclear division | 33/193 | 232/11468 | 0.000 | 4.00e-18 | 0.000 | 33 | CDCA8/CDC20/KIF23/CENPE/MYBL2/CCNB2/... |
| GO:0000280 | nuclear division | 35/193 | 316/11468 | 0.000 | 4.50e-16 | 0.000 | 35 | CDCA8/CDC20/KIF23/CENPE/MYBL2/CCNB2/... |

Figure 5: El resultat d'anàlisi ORA. GO.

2.1.2 KEGG

Per l'ORA de base de dades KEGG he utilitzat la funció `enrichKEGG()` del paquet `clusterProfiler`.

```
enrichKEGG(gene, organism = "hsa", keyType = "kegg", pvalueCutoff = 0.05,
pAdjustMethod = "BH", universe, minGSSize = 10, maxGSSize = 500,
qvalueCutoff = 0.2, use_internal_data = FALSE)
```

Pathway analysis

- Data input
- GO Analysis
- KEGG Analysis
- ORA
- GSEA
- Bar Plot
- Dot Plot
- Enrichment Plot
- Category-gene-network
- GSEA plot
- KEGG Pathway
- Reactome Analysis

Select P-Value threshold:

☒ 0.1 ☐ 0.05 ☐ 0.01 ☐ 0.001

Select adjustment method

holm

Calculate Result

Figure 6: Configuració d'anàlisi KEGG

Una vegada introduïts els paràmetres i premut el botó **Calculate** apareix el botó **Download .csv** i la taula previsualitzada. Els camps de la taula són els mateixos com en l'anàlisi dels termes GO.

Calculate Result

Download Results as .csv

| ID | Description | GeneRatio | BgRatio | pvalue | p.adjust | qvalue | Count | geneID |
|----------|---------------------|-----------|----------|--------|----------|--------|-------|--|
| hsa04110 | Cell cycle | 11/92 | 124/7841 | 0.000 | 3.48e-05 | 0.000 | 11 | 8318/991/9133/890/983/4085/7272/1111/891/4174/9232 |
| hsa04114 | Oocyte meiosis | 10/92 | 125/7841 | 0.000 | 1.70e-04 | 0.000 | 10 | 991/9133/983/4085/51806/6790/891/9232/3708/5241 |
| hsa04218 | Cellular senescence | 10/92 | 160/7841 | 0.000 | 1.04e-03 | 0.001 | 10 | 2305/4605/9133/890/983/51806/1111/891/776/3708 |

Figure 7: El resultat de l'anàlisi ORA. KEGG.

2.1.3 Reactome

En el cas de Reactome el procediment és similar. La funció usada és `enrichPathway()` del paquet `ReactomePA`:

```
enrichPathway(gene, organism = "human", pvalueCutoff = 0.05,
pAdjustMethod = "BH", qvalueCutoff = 0.2, universe, minGSSize = 10,
maxGSSize = 500, readable = FALSE)
```

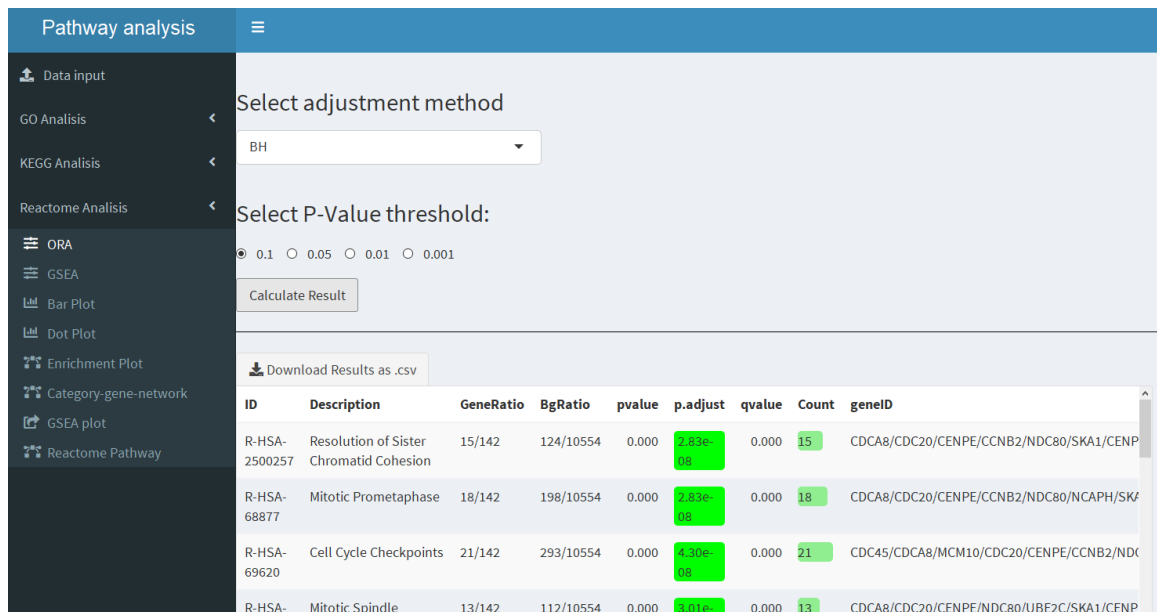


Figure 8: El resultat d'anàlisi ORA. Reactome.

2.2 GSEA

2.2.1 GO

El mètode GSEA per a termes GO es calcula amb la funció `gseGO()` del paquet `clusterProfiler`.

```
gseGO(geneList, ont = "BP", OrgDb, keyType = "ENTREZID",
      exponent = 1, nPerm = 1000, minGSSize = 10, maxGSSize = 500,
      pvalueCutoff = 0.05, pAdjustMethod = "BH", verbose = TRUE,
      seed = FALSE, by = "fgsea")
```

L'usuari pot elegir l'ontologia GO, el *cut-off* del valor P i el mètode d'ajustament.

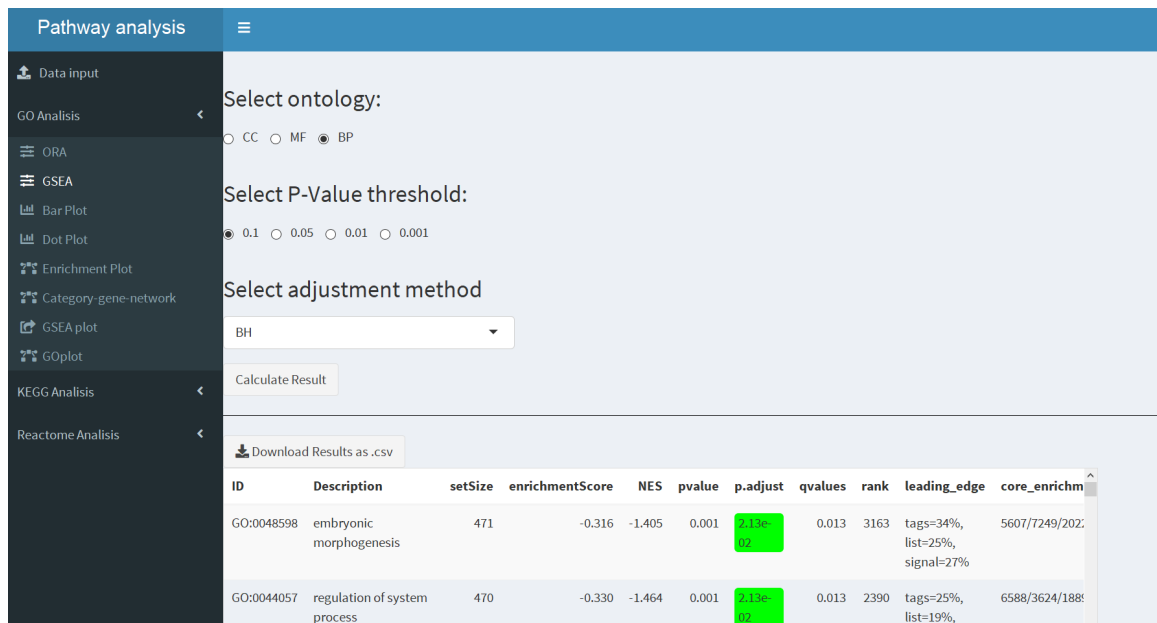


Figure 9: El resultat de l'anàlisi GSEA. GO.

Per entendre l'anàlisi:

- enrichmentScore. Enrichment score per al conjunt dels gens. Amb altres paraules: el grau amb el qual el conjunt dels gens està sobreexpressat a dalt o a baix del llistat ordenat dels gens en les dades d'expressió.
- NES. Normalized enrichment score. La puntuació per al conjunt de gens després de ser normalitzat tenint en compte tots els conjunts de gens analitzats (la seva mida i la seva correlació amb les dades d'expressió). Aquesta puntuació ajuda a comparar els resultats entre els conjunts de gens.
- pvalue. El valor de p nominal.
- p.adjust. El valor de p ajustat.
- leading_edge
 - Tags. El percentatge de les ocurrences de gens del conjunt específic abans (per als ES positius) o després (per als ES negatius) del cim en la puntuació corrent d'enriquement. Aquest valor indica el percentatge dels gens que contribueixen a la puntuació d'enriquement.
 - List. El percentatge dels gens en el llistat ordenat de tots els gens abans o després del pic en la puntuació corrent d'enriquement. Aquest valor ens indica on exactament el pic es produeix.
 - Signal. La fortalesa del senyal d'enriquement que combina els dos valors anteriors.
- rank. La posició del pic en la llista ordenada dels gens. Els conjunts dels gens més interessants assoleixen el seu màxim o bé al principi o al final de la llista ordenada. Vol dir que tenen aquest valor o bé molt baix o bé molt alt.

2.2.2 KEGG

De la mateixa manera es calcula GSEA amb la funció `gseKEGG()` del paquet `clusterProfiler`:

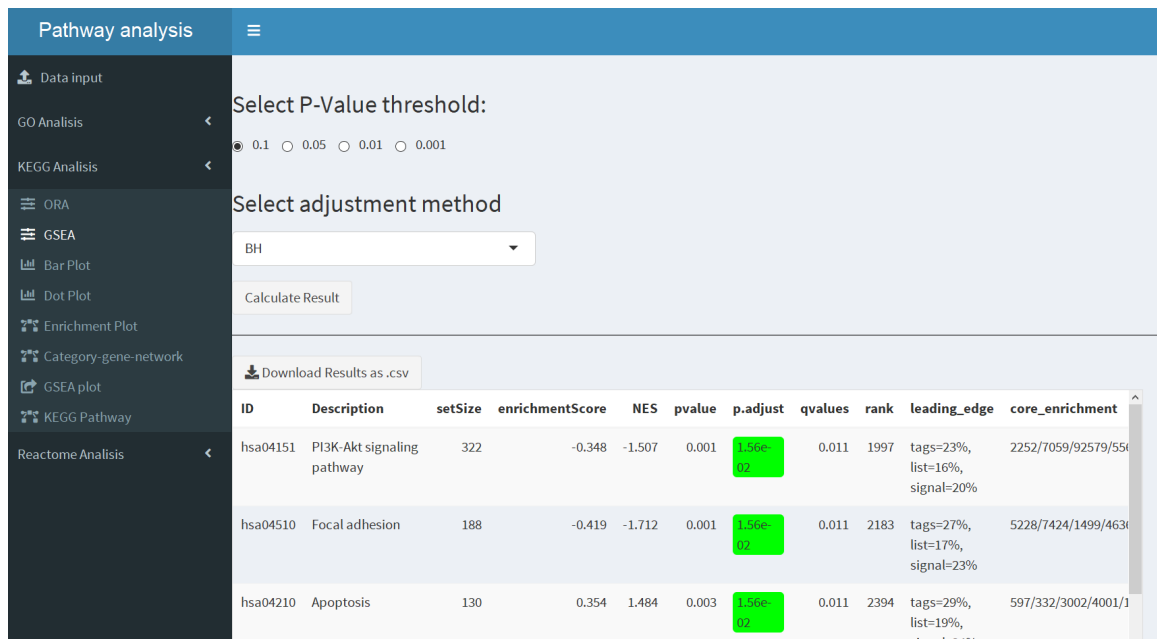


Figure 10: El resultat de l'anàlisi GSEA. KEGG.

2.2.3 Reactome

Per completar l'anàlisi l'usuari pot calcular GSEA per a base de dades Reactome. Com als altres casos utilitzo el paquet `clusterProfiler` i específicament la funció `gsePathway()`

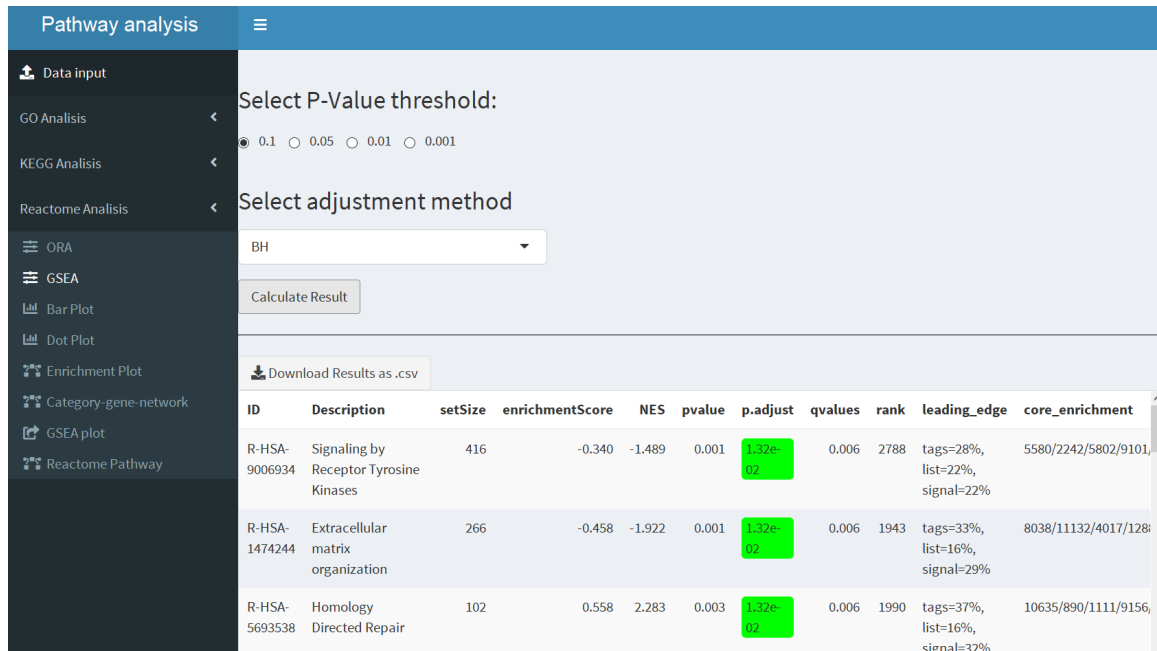


Figure 11: El resultat d'anàlisi GSEA. Reactome.

2.3 Bar-Plots

Els resultats de `enrichGO`, `enrichKEGG` i `enrichPathway` es poden visualitzar amb el gràfic de barres. L'usuari pot elegir el nombre de les categories visualitzades entre 2 i 30. Es dona l'opció per descarregar el gràfic en

format .png.

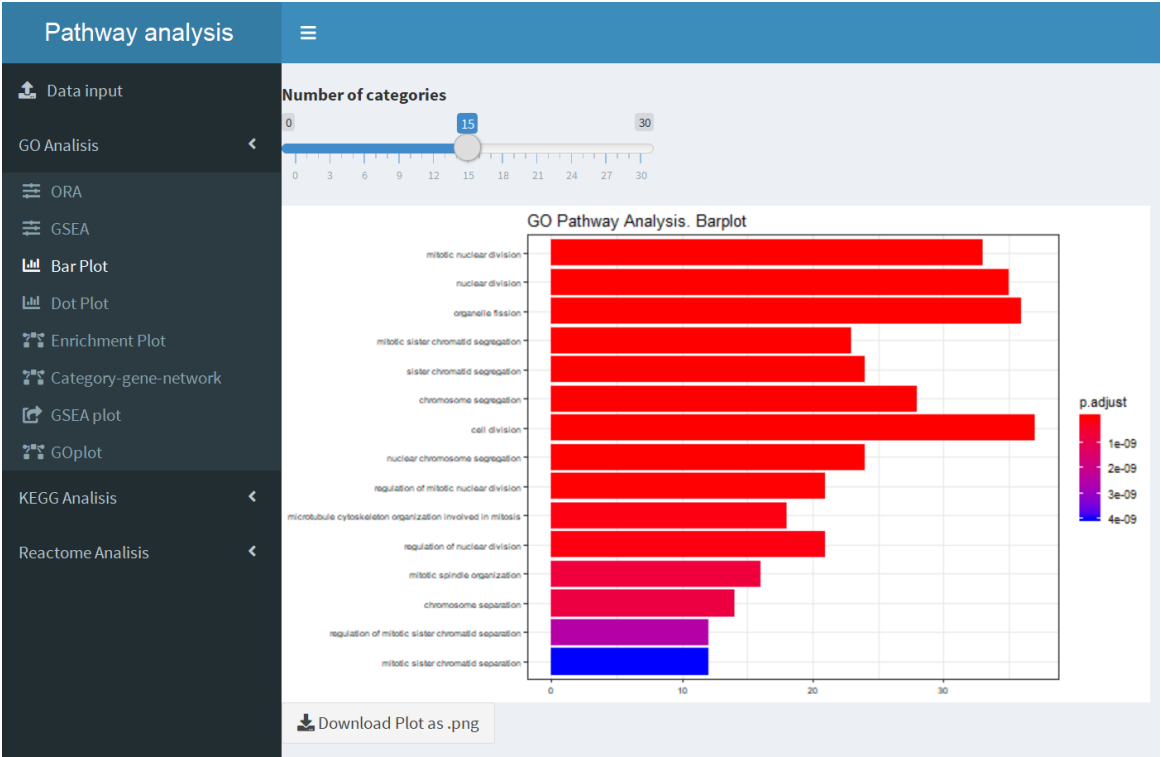


Figure 12: Bar-Plot. GO.

2.4 Dot-Plots

El *dot plot* visualitza addicionalment el *gen ratio*. També aquí l'usuari pot seleccionar el nombre de categories.

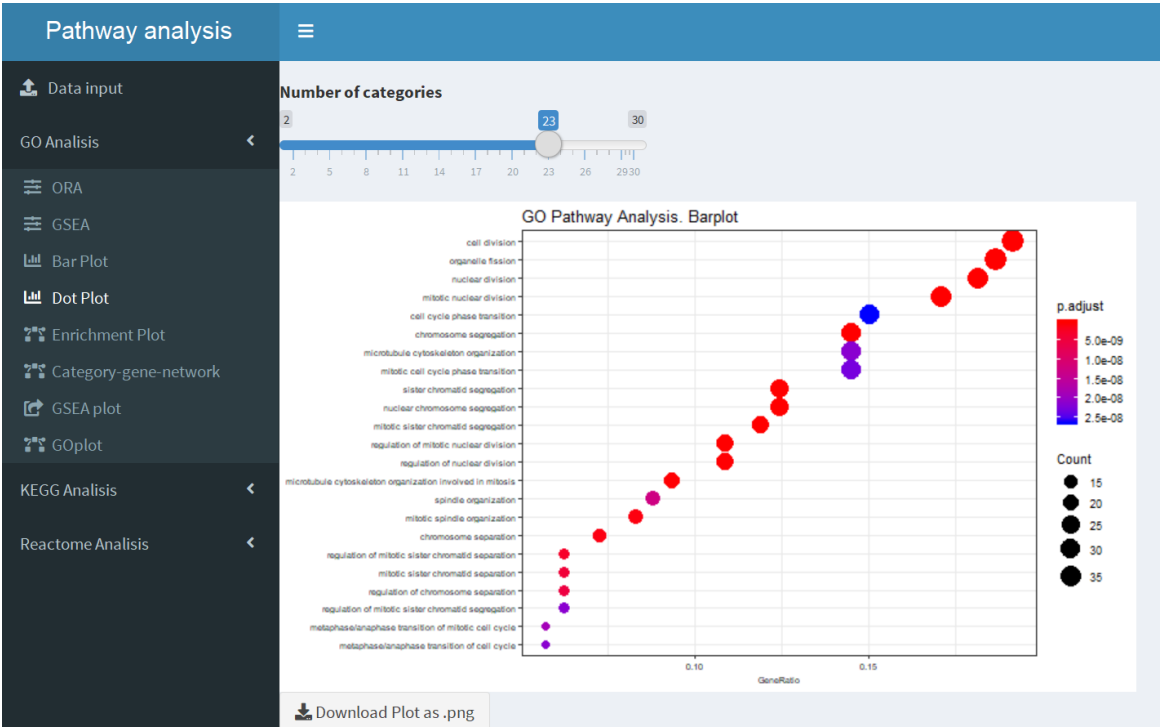


Figure 13: Bar-Plot. GO.

2.5 Enrichment Plots

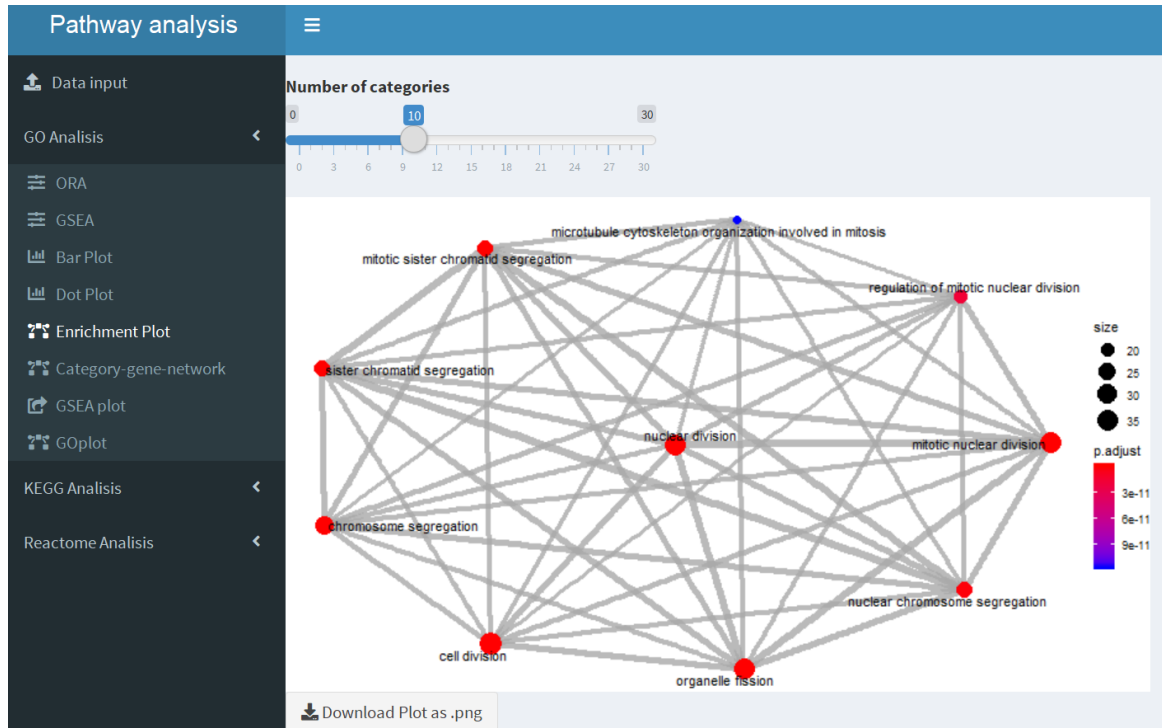


Figure 14: Bar-Plot. GO.

2.6 Category-Gene-Network Plot

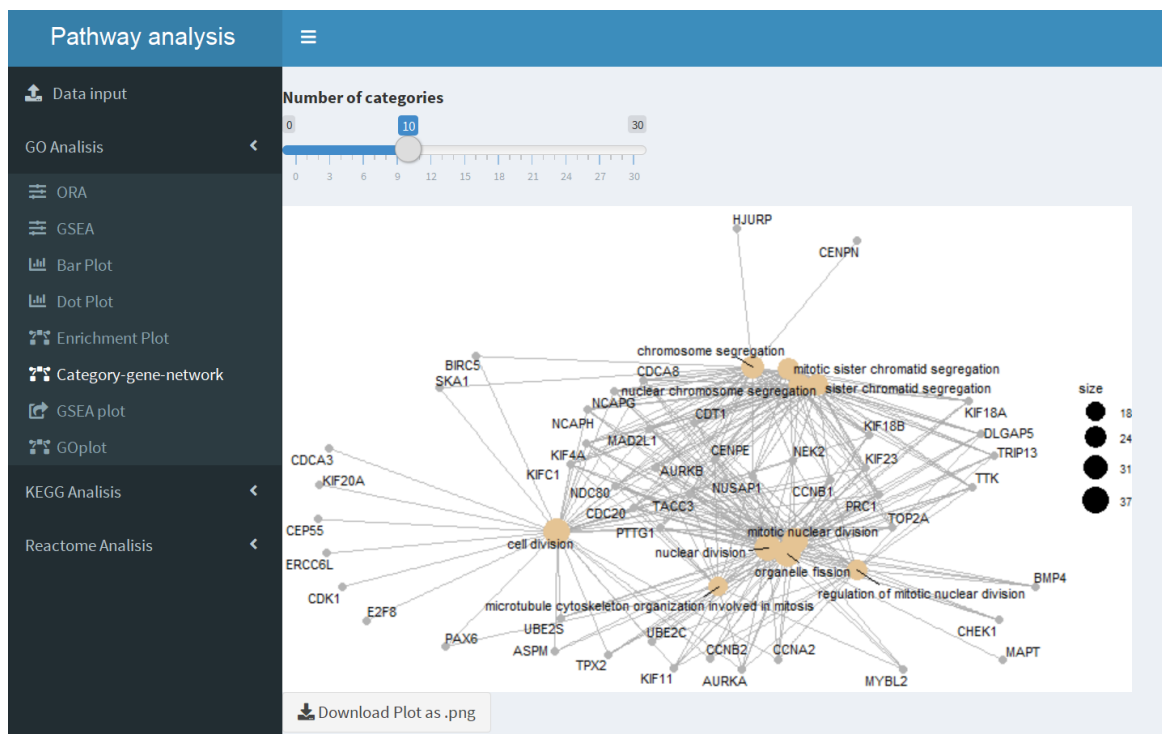


Figure 15: Category-Gene-Network Plot. GO.

2.7 GSEA Plot

L'usuari pot visualitzar una de les categories disponibles via *dropdown list*. El llistat inclou totes les rutes generades durant l'anàlisi GSEA en els apartats *Go Analysis*→*GSEA*; *KEGG*→*GSEA*

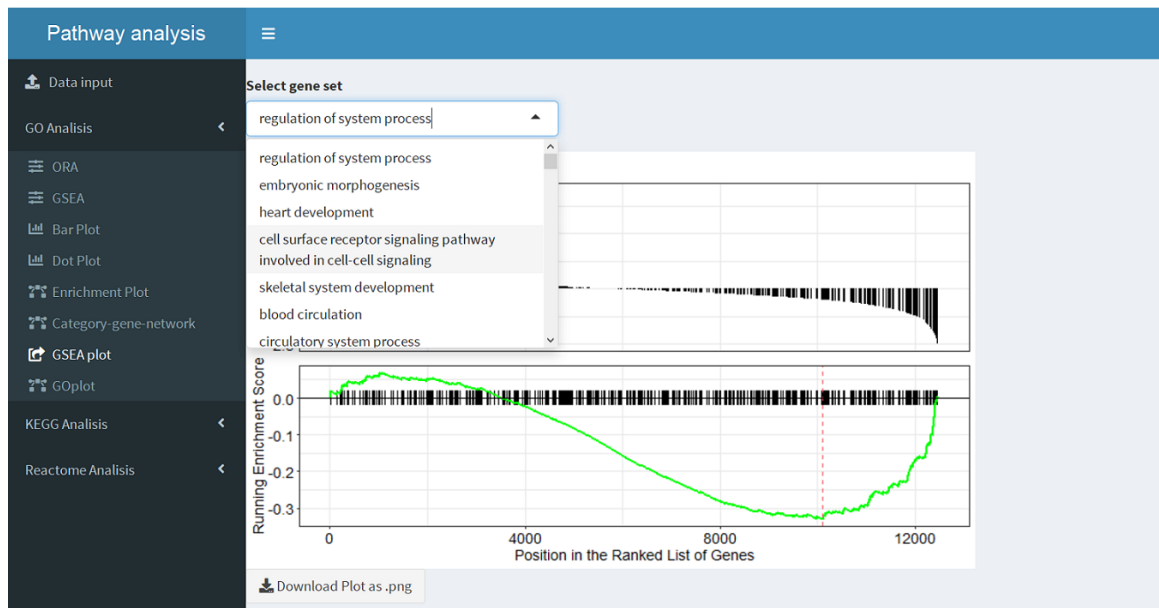


Figure 16: GSEA Plot. GO.

3.2 KEGG Pathway

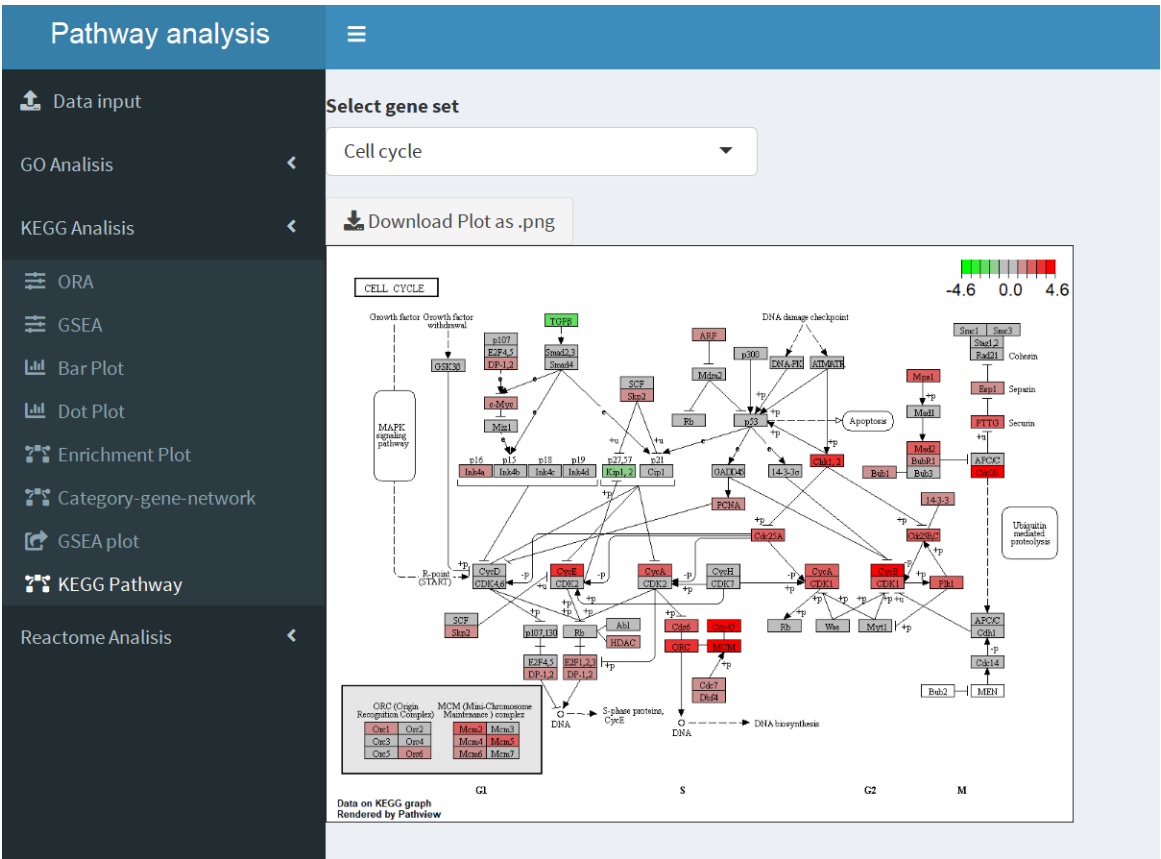


Figure 18: KEGG pathway

3.3 Reactome Pathway

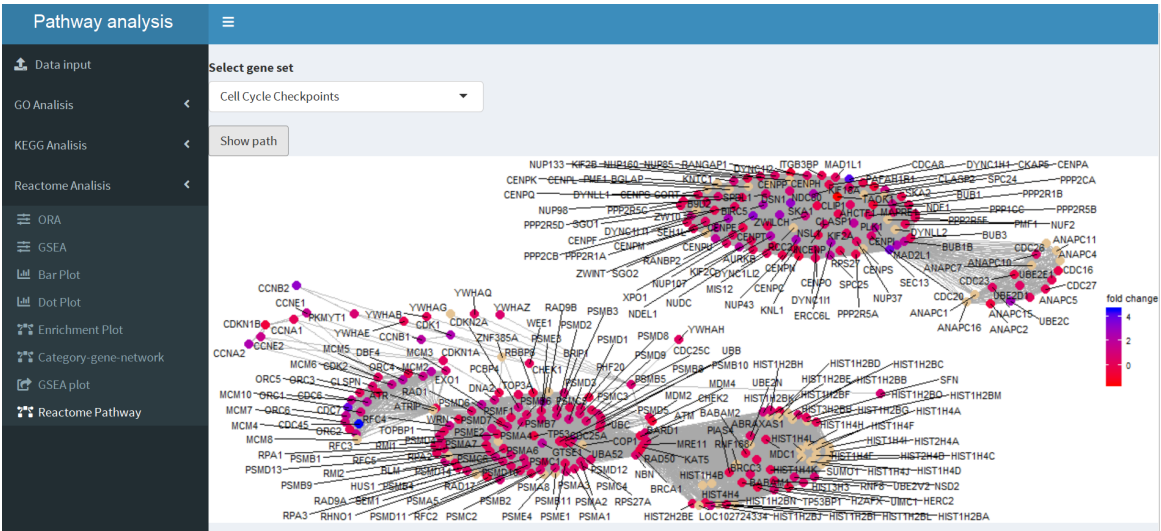


Figure 19: Reactome pathway

4 Validació dels resultats

Per validar l'aplicació he intentat trobar estudis que investiguin les espècies diferents i utilitzin com el tipus d'experiment o bé *Microarrays* o bé *NGS (New Generation Sequencing)*. En la base de dades GEO (Gene Expression Omnibus) he elegit els estudis següents:

| Estudi | GEO ID | Espècie | Tipo d'experiment |
|------------------------|-----------|--------------|-------------------|
| [Schmidt et al., 2008] | GSE11121 | Homo sapiens | Microarrays |
| [Li et al., 2017] | GSE100924 | Mus musculus | Microarrays |

L'aplicació necessita un arxiu amb les IDs de Entrez i els LogChanges. Malauradament GEO no disposa d'aquestes dades. Només té les dades en format .cel. Per tant, primer s'hauria fer els passos d'anàlisi de dades d'expressió per derivar les dades que m'interessen. Aquests passos inclouen: normalització d'una banda i calculació del model per derivar els LogChanges d'altra banda.

Les dades d'estudi [Li et al., 2017] són ja preprocessades per Ricardo Gonzalo Sanz i Sanchez Pla i estan disponibles a github. De la carpeta *results* he agafat les taules *topAnnotated_KOvsWT_COLD.csv* i *topAnnotated_KOvsWT_RT.csv*. Sanz i Pla utilitzen el paquet **ReactomePA** per a l'anàlisi d'enriquement. Repeteixo doncs el seu anàlisi utilitzant l'aplicació.

1. Ellegeixo la espècie *Mus musculus* per a GO, KEGG i Reactome.

The screenshot shows the 'Pathway analysis' web application interface. On the left is a dark sidebar with a 'Data input' section containing links for 'GO Analysis', 'KEGG Analysis', and 'Reactome Analysis'. The main content area is light blue and divided into three sections: 'GO', 'Reactome', and 'KEGG'. Each section has a 'Select Species:' dropdown menu currently set to 'Mouse'. The 'KEGG' section also includes a text input for 'Enter Search Term for Species' with 'mus' entered, and a 'Select KEGG Species' dropdown set to 'Mus musculus'. Below these are two file upload sections: 'File with all genes' and 'File with selected genes'. Each has a 'Browse...' button and a file name (e.g., 'topAnnotated_KOvsWT_COLD_ur' and 'topAnnotated_KOvsWT_COLD_ge') with an 'Upload complete' button. A note at the bottom states: 'Note: The file should contain two columns, one for gene ID (no duplicated ID allowed) and another one for fold change.'

Figure 20: Selecció d'espècie

L'output a baix indica que s'ha pujat el total de 5995 gens. Per a l'arxiu dels gens seleccionats l'aplicació diu que s'han pujat 769 gens.

Note: The file should contain two columns, one for gene ID (no duplicated ID allowed) and another one for fold change.

You uploaded: 5995 genes

First 10 entries

| Entrez ID | FoldChange |
|-----------|------------|
| 108664 | -0.420 |
| 319263 | 0.049 |
| 59014 | -0.143 |
| 109294 | 0.114 |
| 320492 | -1.454 |
| 98711 | 0.072 |
| 17087 | -0.653 |
| 75712 | -0.384 |
| 14859 | -0.378 |
| 27993 | -0.113 |

You selected: 769 genes

First 10 entries

| Entrez ID | FoldChange |
|-----------|------------|
| 320492 | -1.454 |
| 50785 | 0.743 |

Figure 21: Selecció d'espècie

2. Clico en l'apartat *Reactome Analysis*→*ORA*. Selecciono com a mètode d'ajustament *BH* i el cut-off del valor de P ajustat 0.05. Clico a *Calculate results*

Pathway analysis

Select adjustment method

BH

Select P-Value threshold:

0.1 0.05 0.01 0.001

Calculate Result

Download Results as .csv

| ID | Description | GeneRatio | BgRatio | pvalue | p-adjust | qvalue | Count | geneID |
|---------------|--|-----------|----------|--------|----------|--------|-------|---|
| R-MMU-75105 | Fatty acyl-CoA biosynthesis | 8/342 | 36/9114 | 0 | 0.017 | 0.017 | 8 | Elov17/Elov12/Ppt2/Elov3/Acsl5/Hsd17b12/Cbr4/Acsl4 |
| R-MMU-8978868 | Fatty acid metabolism | 20/342 | 196/9114 | 0 | 0.017 | 0.017 | 20 | Alox5/Slc22a21/Acot1/Acot8/Elov17/Ec13/Elov12/Ppt2/Cp |
| R-MMU-75876 | Synthesis of very long-chain fatty acyl-CoAs | 6/342 | 22/9114 | 0 | 0.031 | 0.031 | 6 | Elov17/Elov12/Elov3/Acsl5/Hsd17b12/Acsl4 |

Figure 22: Resultat d'anàlisi ORA de Reactome

Observem que els gens mostrats són els mateixos esmentats per Sanz i Pla.

3. Visualització del resultat ORA

- Selecciono *Reactome Analysis*→*Bar Plot*

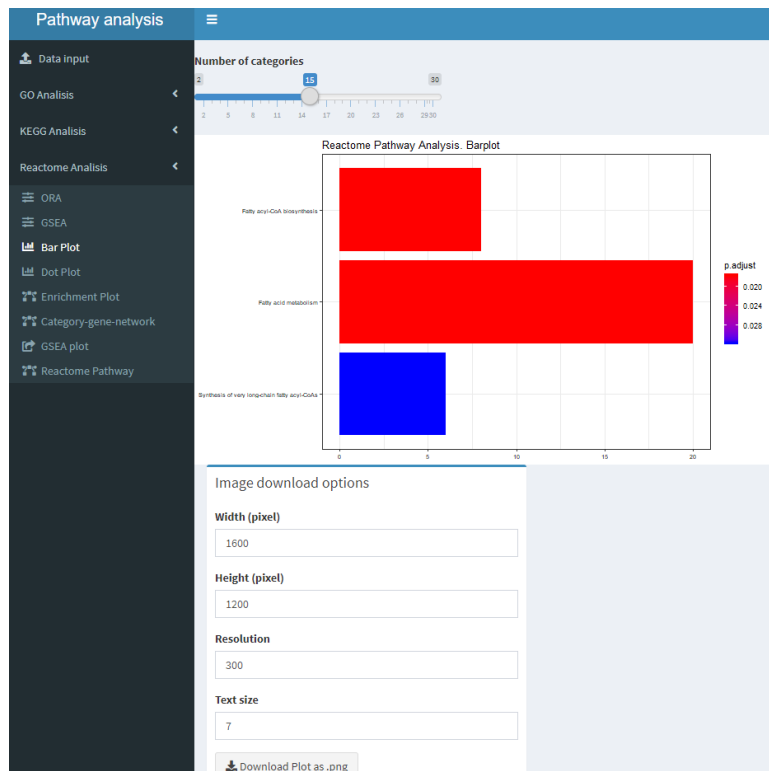


Figure 23: Gràfic de barres

- Selecciono *Reactome Analysis*→*Dot Plot*

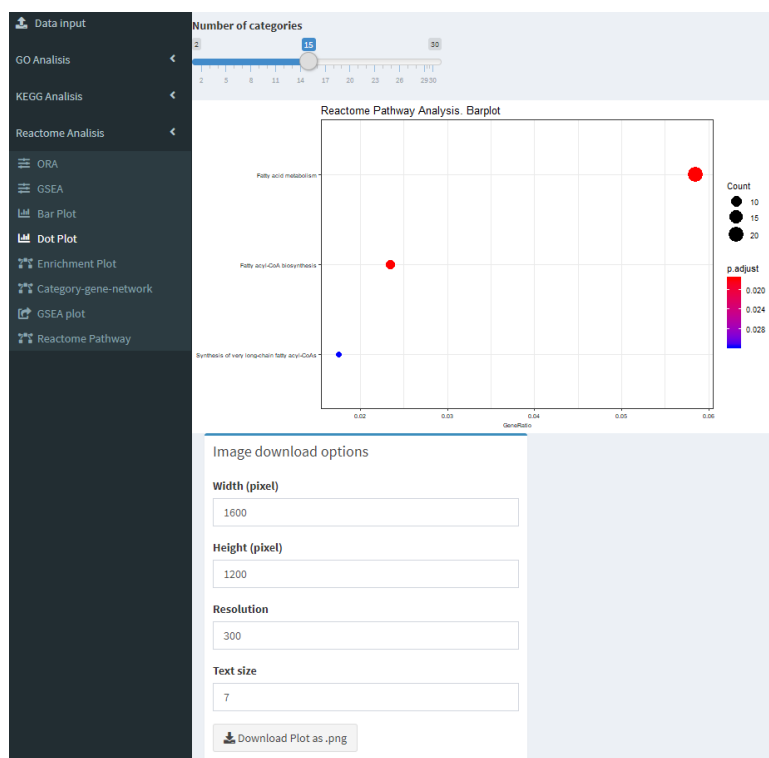


Figure 24: Gràfic de punts

- Selecciono *Reactome Analysis*→*Enrichment Map Plot*

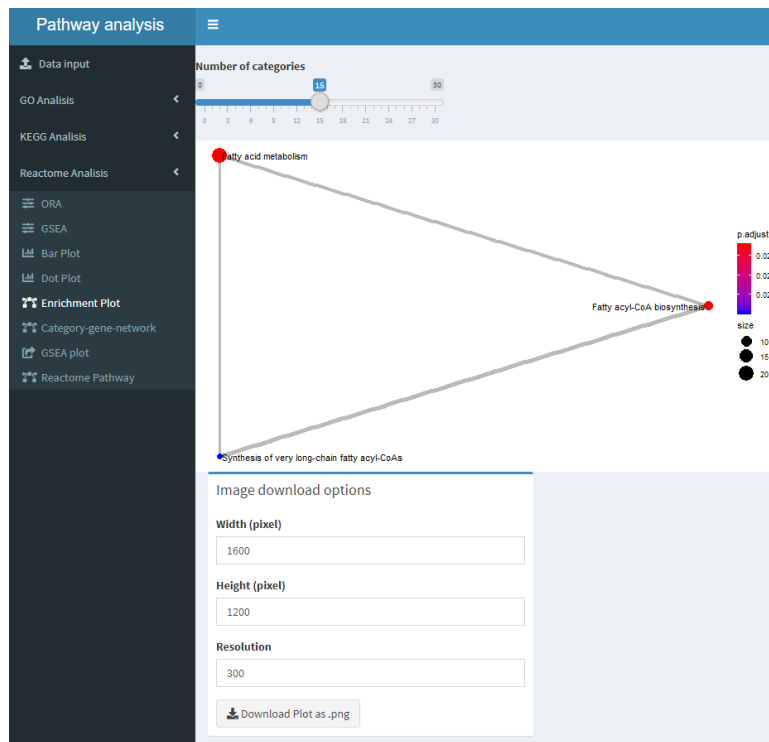


Figure 25: Mapa d'enriquement

- Selecciono *Reactome Analysis*→*Category Gene Network*

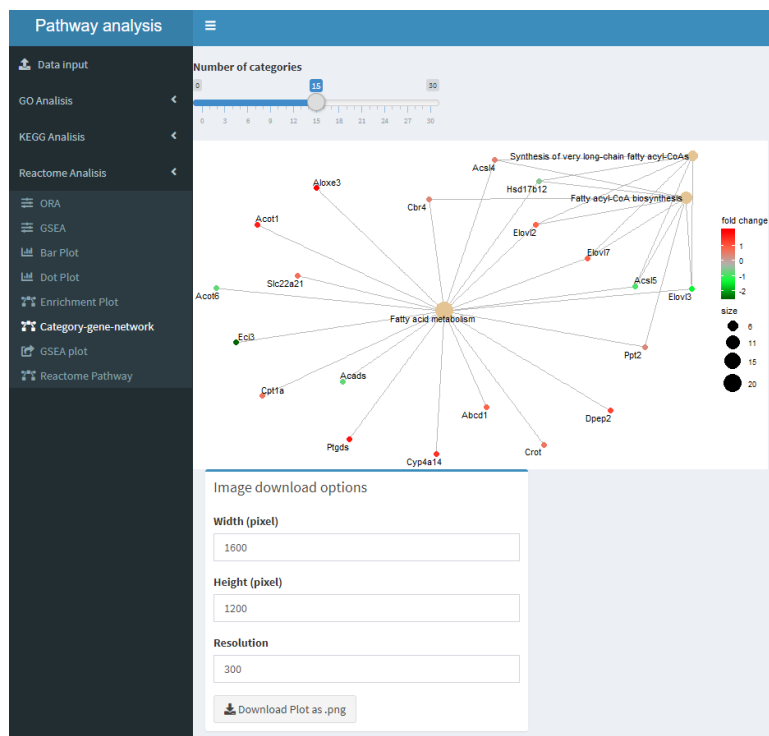


Figure 26: Red de les categories i gens

- Selecciono *Reactome Analysis*→*Reactome Pathway*

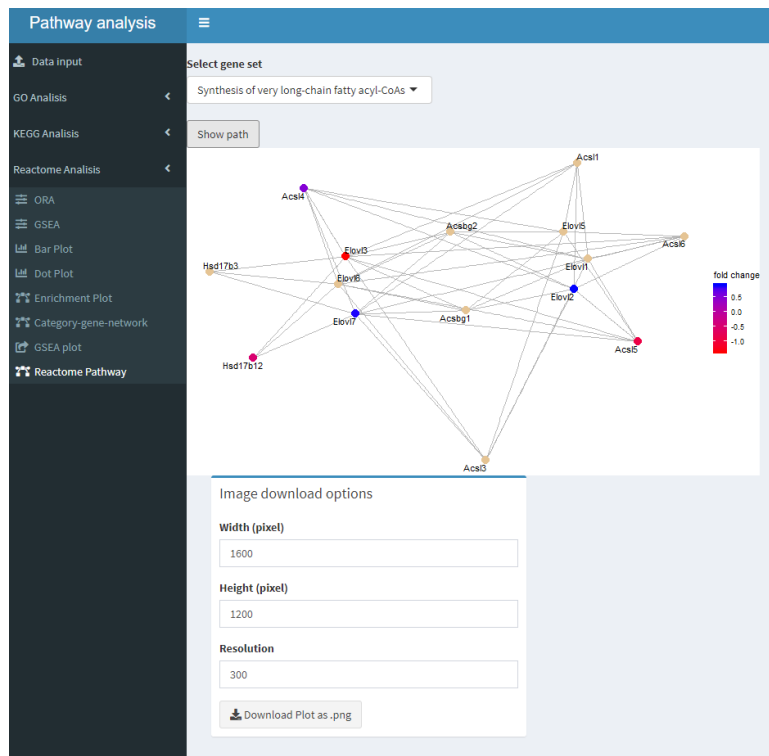


Figure 27: Rutes Reactome

Addicionalment a l'anàlisi ORA podem fer mitjançant l'aplicació l'anàlisi GSEA per les rutes de Reactome. Per fer-ho:

1. Clico en l'apartat *Reactome Analysis* → *GSEA*. Selecciono com a mètode d'ajustament *BH* i el cut-off del valor de P ajustat 0.05. Clico a *Calculate results*

Amb el valor de P de 0.05 l'anàlisi no troba cap ruta enriquida.

2. Augmento el Cut-Off del valor de P a 0.1

Amb el Cut-Off més alt l'aplicació retorna un llistat de gens.

| Pathway analysis | | | | | | | | | | |
|--|---|---------|-----------------|--------|--------|----------|---------|------|--------------------------------|---------------------|
| Select P-Value threshold: | | | | | | | | | | |
| <input checked="" type="radio"/> 0.1 <input type="radio"/> 0.05 <input type="radio"/> 0.01 <input type="radio"/> 0.001 | | | | | | | | | | |
| Select adjustment method | | | | | | | | | | |
| BH | | | | | | | | | | |
| Calculate Result | | | | | | | | | | |
| Download Results as .csv | | | | | | | | | | |
| ID | Description | setSize | enrichmentScore | NES | pvalue | p.adjust | qvalues | rank | leading_edge | core_enrichment |
| R-MMU-69242 | S Phase | 43 | -0.528 | -1.851 | 0.002 | 0.002 | 0.032 | 1968 | tags=53%, list=33%, signal=36% | 22190/18971/109145 |
| R-MMU-8852276 | The role of GTSE1 in G2/M progression after G2 checkpoint | 27 | -0.592 | -1.871 | 0.002 | 0.002 | 0.032 | 614 | tags=37%, list=10%, signal=33% | 15516/16912/23996/2 |
| R-MMU-174143 | APC/C-mediated degradation of cell cycle proteins | 26 | -0.669 | -2.100 | 0.002 | 0.002 | 0.032 | 703 | tags=42%, list=12%, signal=38% | 66413/12237/16912/2 |
| R-MMU-453276 | Regulation of mitotic cell cycle | 26 | -0.669 | -2.100 | 0.002 | 0.002 | 0.032 | 703 | tags=42%, list=12%, signal=38% | 66413/12237/16912/2 |

Figure 28: Anàlisi GSEA

3. Per obtenir els gràfics GSEA anem a *Reactome Analysis*→*GSEA plot*

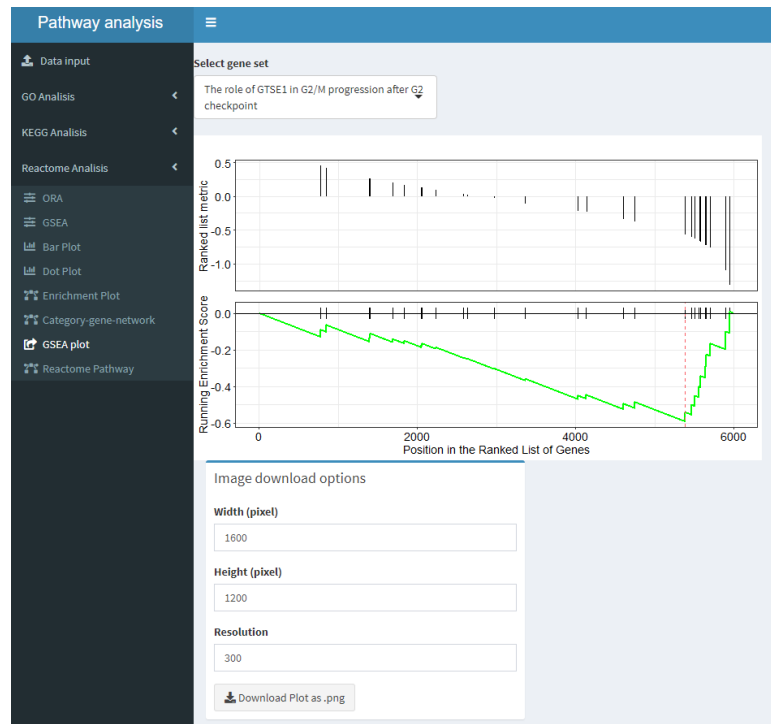


Figure 29: Gràfic GSEA

També podem fer l'anàlisi de KEGG. El resultat de KEGG és similar a l'anàlisi de Reactome. L'aplicació permet però generar les rutes KEGG. Per obtenir-les:

1. Clico en l'apartat *KEGG Analysis*→*ORA*. Selecciono com a mètode d'ajustament *BH* i el cut-off del valor de P ajustat 0.05. Clico a *Calculate results*

| ID | Description | GeneRatio | BgRatio | pvalue | p.adjust | qvalue | Count | geneID |
|----------|---|-----------|---------|--------|----------|--------|-------|--|
| mmu01212 | Fatty acid metabolism | 10/285 | 57/8601 | 0.000 | 0.004 | 0.004 | 10 | 74559/54326/54397/12894/12686/433256/56348/11409 |
| mmu00062 | Fatty acid elongation | 7/285 | 29/8601 | 0.000 | 0.004 | 0.004 | 7 | 26897/74559/54326/54397/12686/56348/83603 |
| mmu01040 | Biosynthesis of unsaturated fatty acids | 6/285 | 32/8601 | 0.001 | 0.044 | 0.044 | 6 | 26897/74559/54326/12686/56348/83603 |

Figure 30: Anàlisi ORA de KEGG

2. Anem a *KEGG*→*KEGG Pathway*

5 Activitats no previstes

Treballant en el projecte he notat la necessitat de fer tot el procés més segur. El moment clau era quan no he pogut trobar l'USB on he guardat el meu projecte. Ho tenia en USB perquè en treballava dels molts ordinadors diferents: de casa, de feina, en portàtil quan era de viatge. Per tan he decidit guardar tot el projecte en github.com. He creat un repositori al qual puc accedir des d'ordinadors diferents.

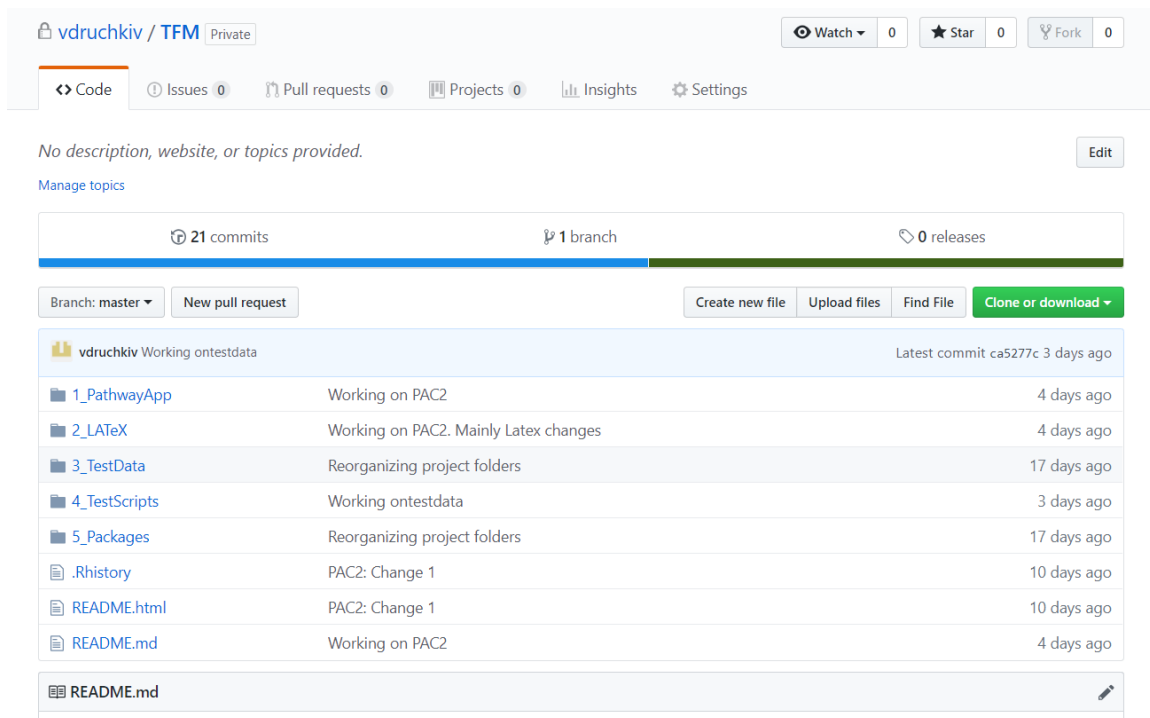


Figure 33: Github repositori del TFM

References

- [Li et al., 2017] Li, S., Mi, L., Yu, L., Yu, Q., Liu, T., Wang, G.-X., Zhao, X.-Y., Wu, J., and Lin, J. D. (2017). Zbtb7b engages the long noncoding rna blnc1 to drive brown and beige fat development and thermogenesis. *Proceedings of the National Academy of Sciences*, 114(34):E7111–E7120.
- [Schmidt et al., 2008] Schmidt, M., Böhm, D., Von Törne, C., Steiner, E., Puhl, A., Pilch, H., Lehr, H.-A., Hengstler, J. G., Kölbl, H., and Gehrmann, M. (2008). The humoral immune system has a key prognostic impact in node-negative breast cancer. *Cancer research*, 68(13):5405–5413.