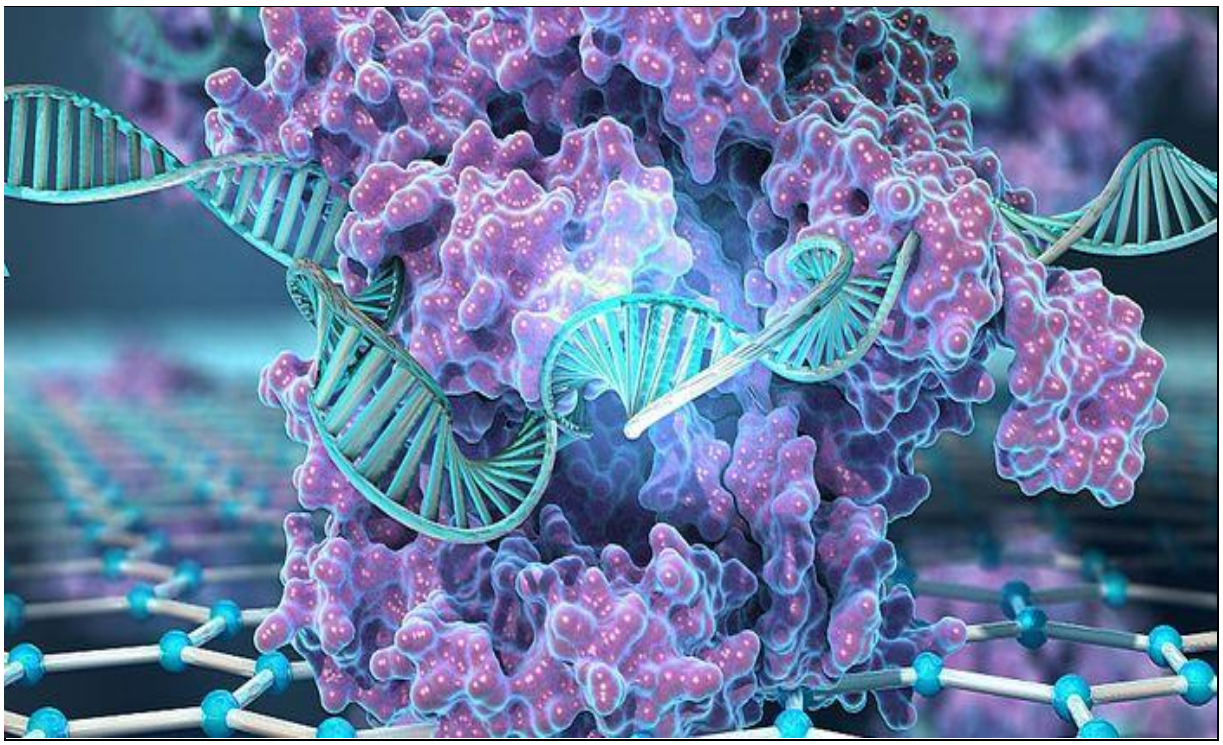


Transcriptomic Data analysis

OF MICROGLIAL EXTRACELLULAR TRAPS

Dr. Sushmitha Jha | Design Credit project | Semester I – 2023-24

OVERVIEW

The experimental sets were taken for RNA-seq and the resulting transcriptomic data was taken up by us. As per the needs and instructions from the mentors, we have started analyzing the transcriptomic data by considering the pre-processed, pre-cleaned and readily available dataset of 'differential expression of genes', and data set of 'inflammation genes.'

- The differential expression data set: https://docs.google.com/spreadsheets/d/1UOzf8j82OhojXm7ytCbfvSAWasqb_dDB/edit?usp=drive_link&ouid=110282695971381908203&rtpof=true&sd=true
- The gene expression counts data sets (T vs C): https://docs.google.com/spreadsheets/d/1U3slyfMXFbPol4Y2koSLAE6khW8NAubF/edit?usp=drive_link&ouid=110282695971381908203&rtpof=true&sd=true
- The complete folder of transcriptomic data analysis: https://drive.google.com/drive/folders/1jqlGw6pujomTvsS5LoZXTBPYxoK_Ur_2?usp=drive_link

WORK FLOW

We have expressed the above-mentioned dataset by analyzing them into multiple components and are:

1. **Differential Expression and Error bars**
2. **KEGG pathway Analysis**
3. **Gene ontology enrichment analysis**
4. **Data filtration**
5. **Heat maps**
6. **Gene functional annotations**

❖ Differential Expression

The total number of genes is around 57000 that have their Fold change values precisely ranging from -11 to +11. Based on the fold changes values, we have taken up the data into multiple sets as

1. <= -10 and >= +10
   The total number of genes in this range were 19.
2. <= -8 and >= +8
   The total number of genes in this range were 361.
3. <= -5 and >= +5
   The total number of genes in this range were 3212.

We had performed the differential expression analysis through Volcano plotting using **Python programming**. The drive folder link for all the plots: https://drive.google.com/drive/folders/1rZ1dkbPtpbqheYKUv_DwbOkrLy5hJySz?usp=drive_link

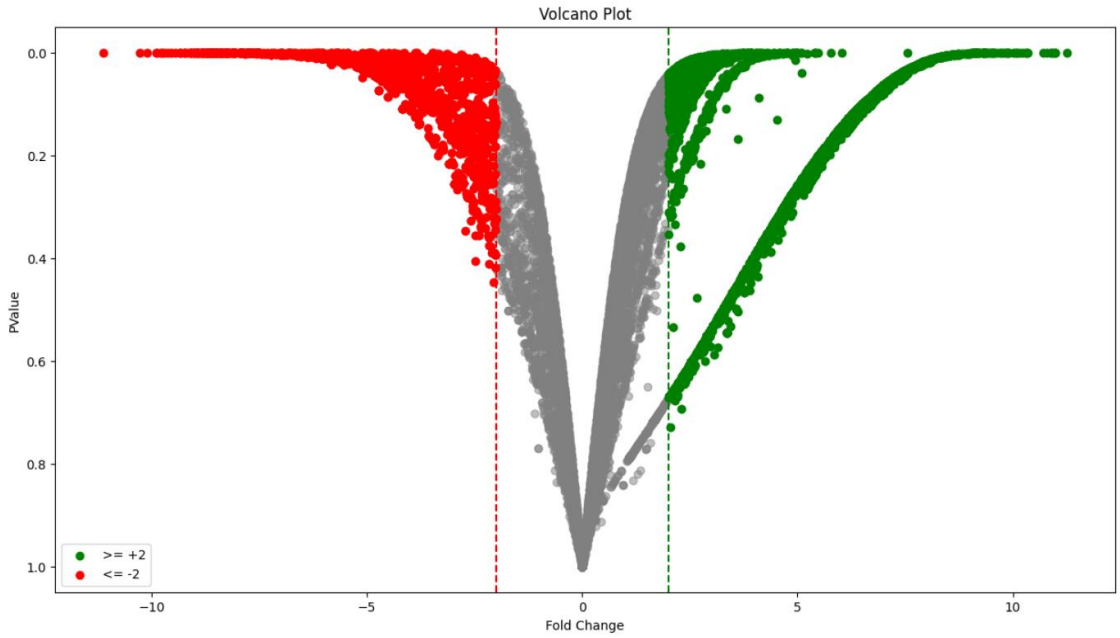


Fig: Volcano plot for differentially expressed genes in the fold change **range <= -2 and >=+2**

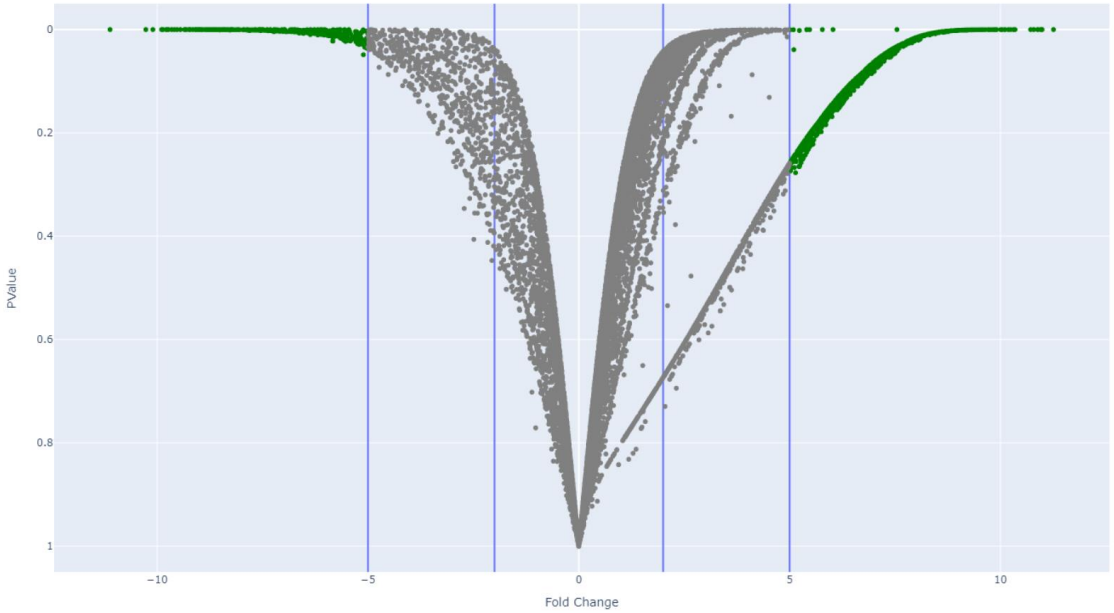


Fig: Volcano plot for differentially expressed genes in the fold change **range <= -5 and >=+5**

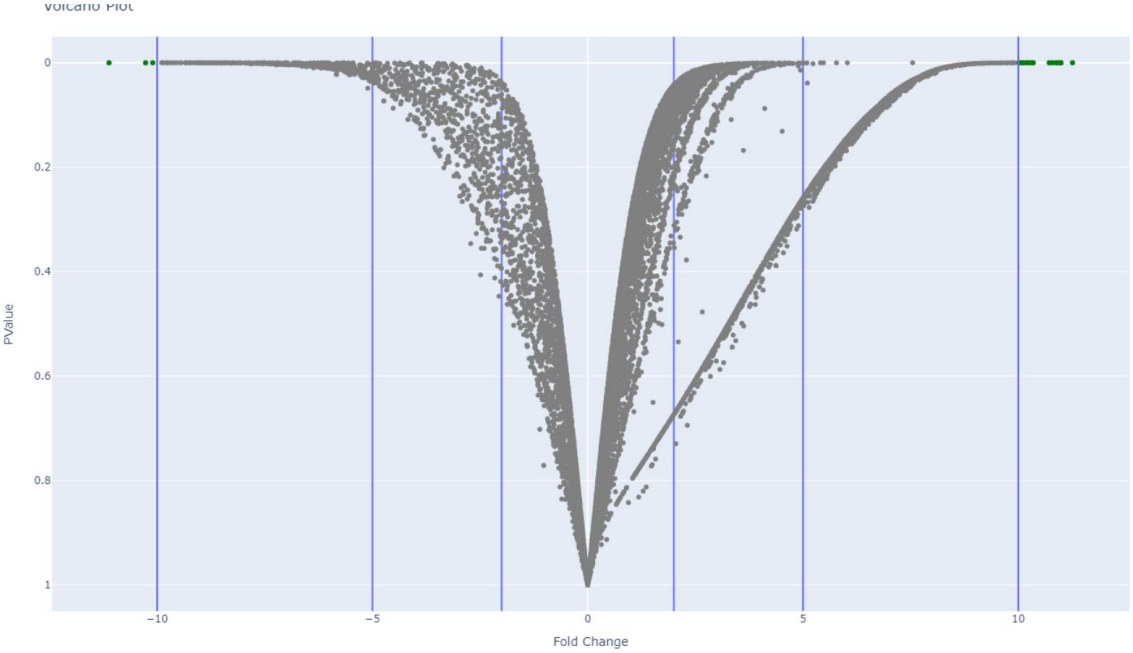


Fig: Volcano plot for differentially expressed genes in the fold change **range <= -10 and >=+10**

❖ Error bars

We have made error plots mainly for all the differentially expressed genes that has fold change values <= -5 and >= +5. The number of genes was so large and so we made the plots breaking the whole set into small ranges and expressed them as their up or down regulation. One of them is given here in the figure. We have used **python programming** for making these plots.

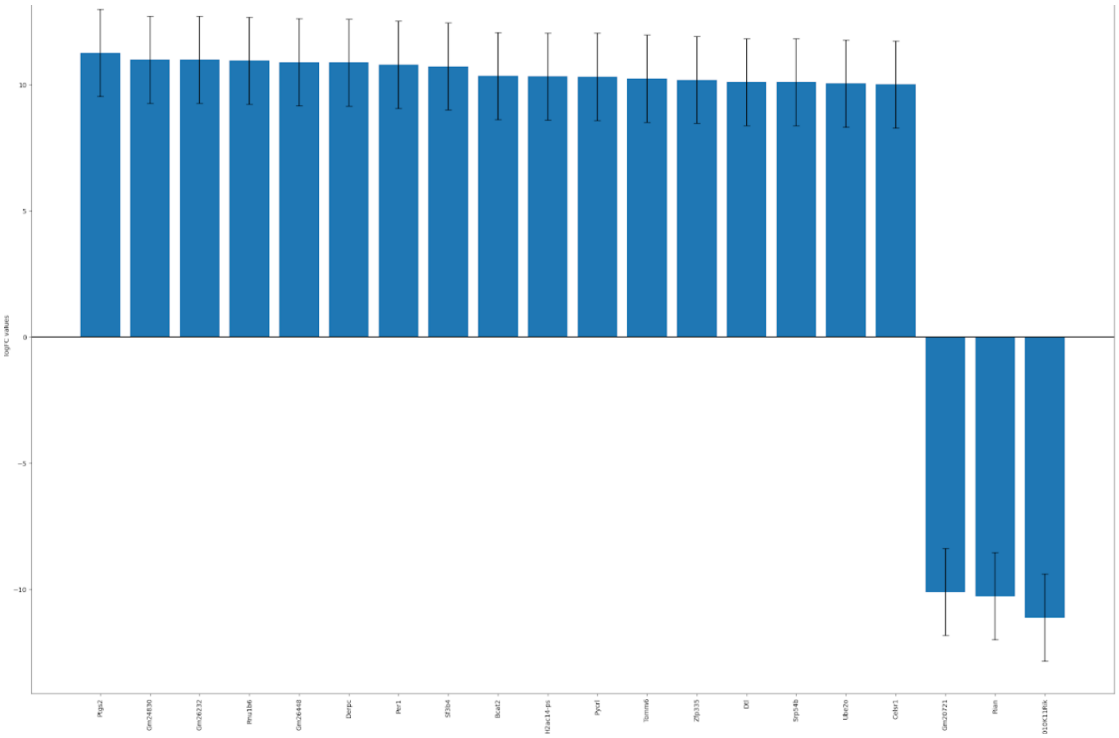


Fig: Regulation expression bar plot for differentially expressed genes and in the fold change **range <= -10 and >= +10.**

The drive folder link for all the plots:
https://drive.google.com/drive/folders/1ePv4UKkIPAT-aaU3HpGIcG_FMELyXJtm?usp=drive_link

❖ KEGG pathway analysis

We have obtained all the KEGG pathways using **SR plots** (the software). Firstly, we have filtered the whole data set of differential expression into subsets of ranges <= -5 and >= +5 , <= -8 and >= +8 , and <= -10 and >= +10. Once the data was filtered, by considering the individual subsets one at a time, we have performed **KEGG pathways analysis** and **Gene ontology enrichment analysis** obtained the possible biological pathways.

We have also taken up the genes in **the inflammation dataset** all at a time and performed both the analysis.

The drive folder link for the same is:
https://drive.google.com/drive/folders/1WdDCxARyUzxnFblXUZFSPQYtHoroACm-?usp=drive_link

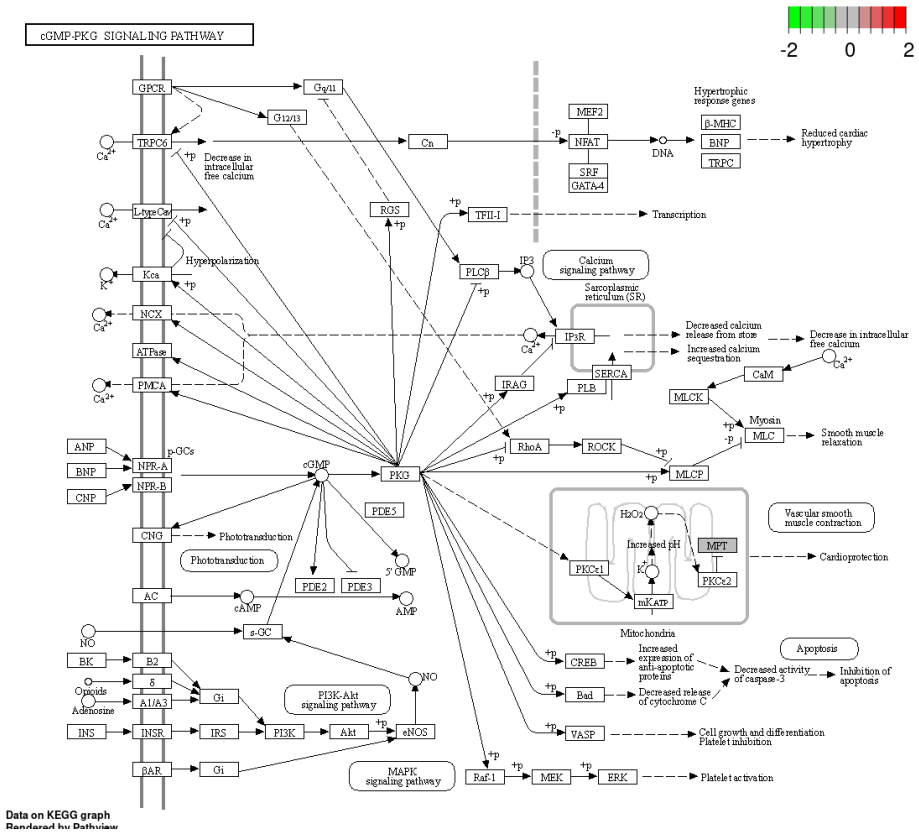


Fig: cGMP-PKG signaling pathway of differentially expressed genes (+-5)

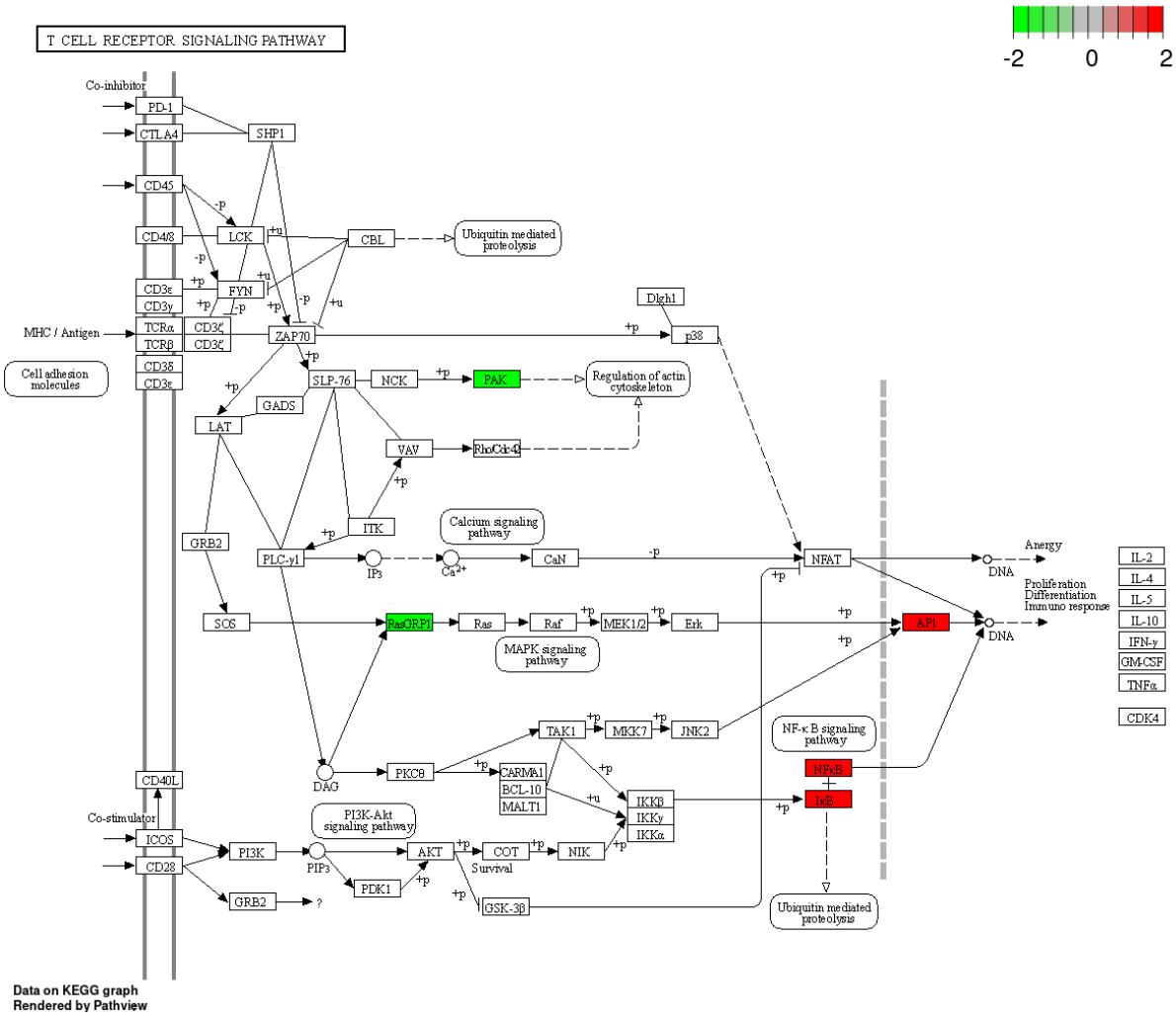


Fig: T cell receptor signaling pathway of differentially expressed genes (+-8)

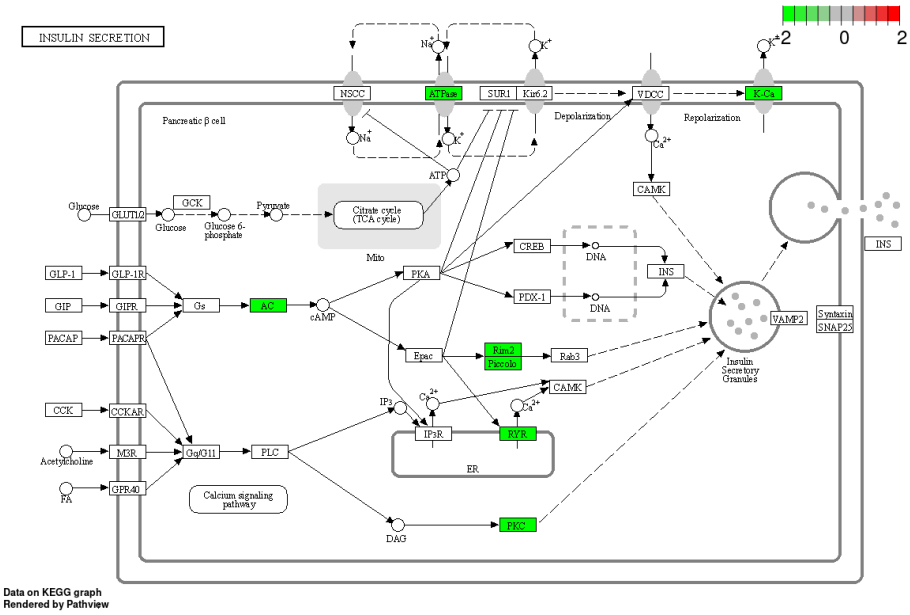


Fig: Insulin secretion pathway of differentially expressed genes (+-8)

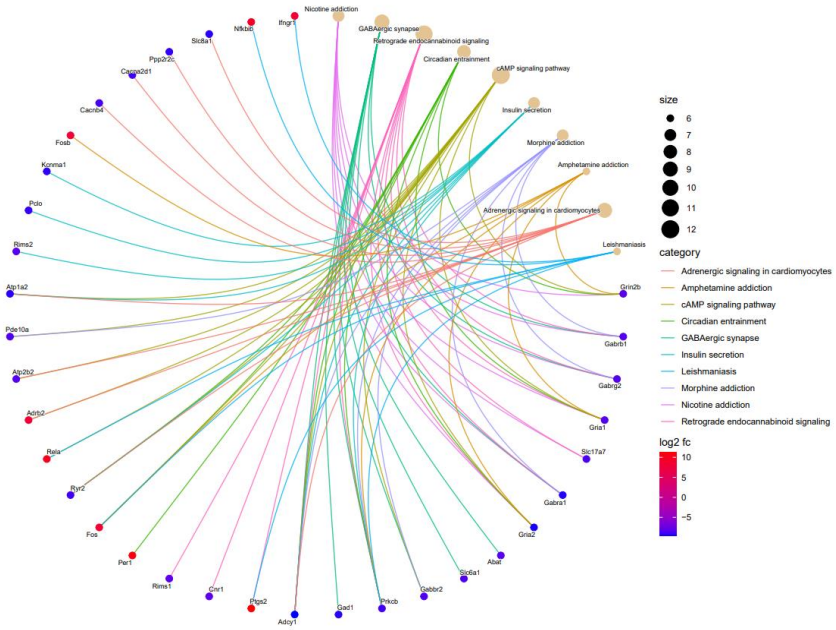


Fig: Network representation of pathways and involved genes (+-8)

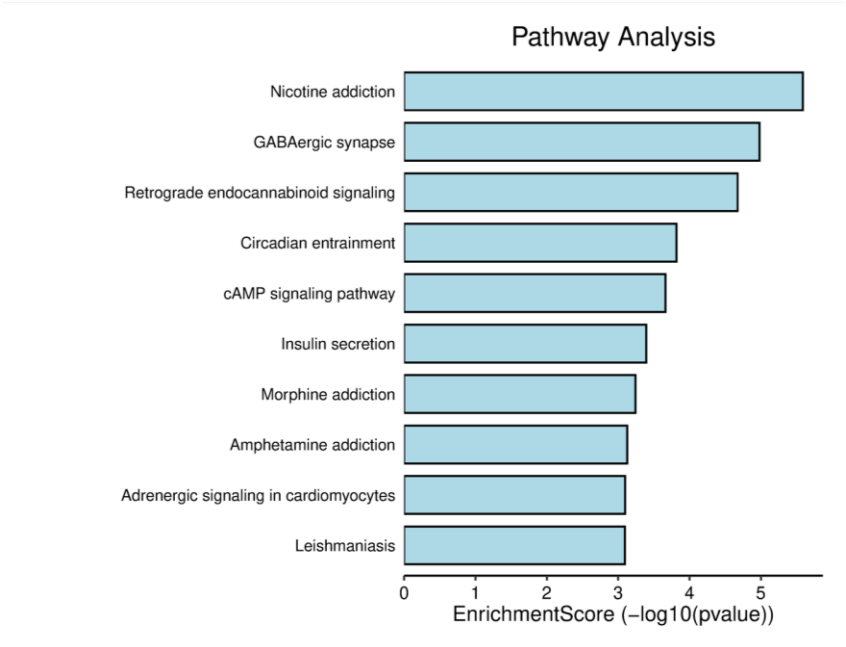


Fig: Enrichment score for the set **KEGG pathways in bar plot (+-8)**

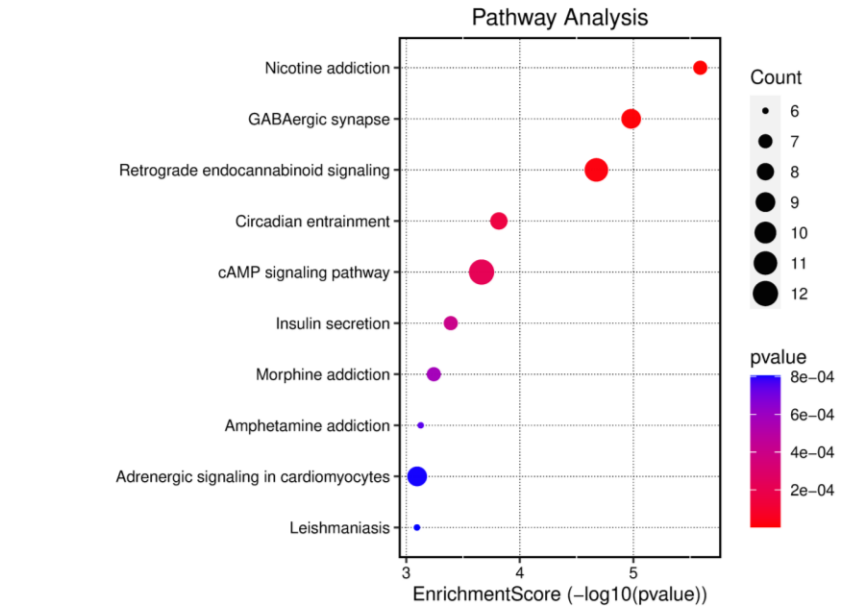


Fig: Enrichment score for the **KEGG pathways in dot plot (+-8)**

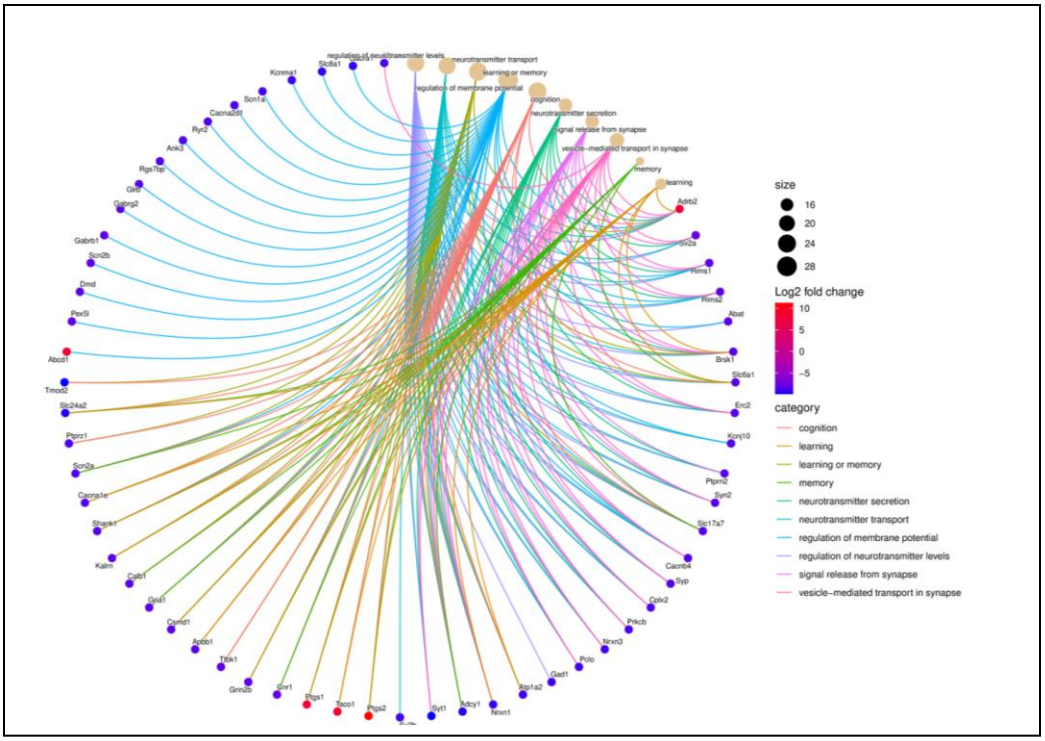


Fig: Network representation of **biological processes (+-8)**

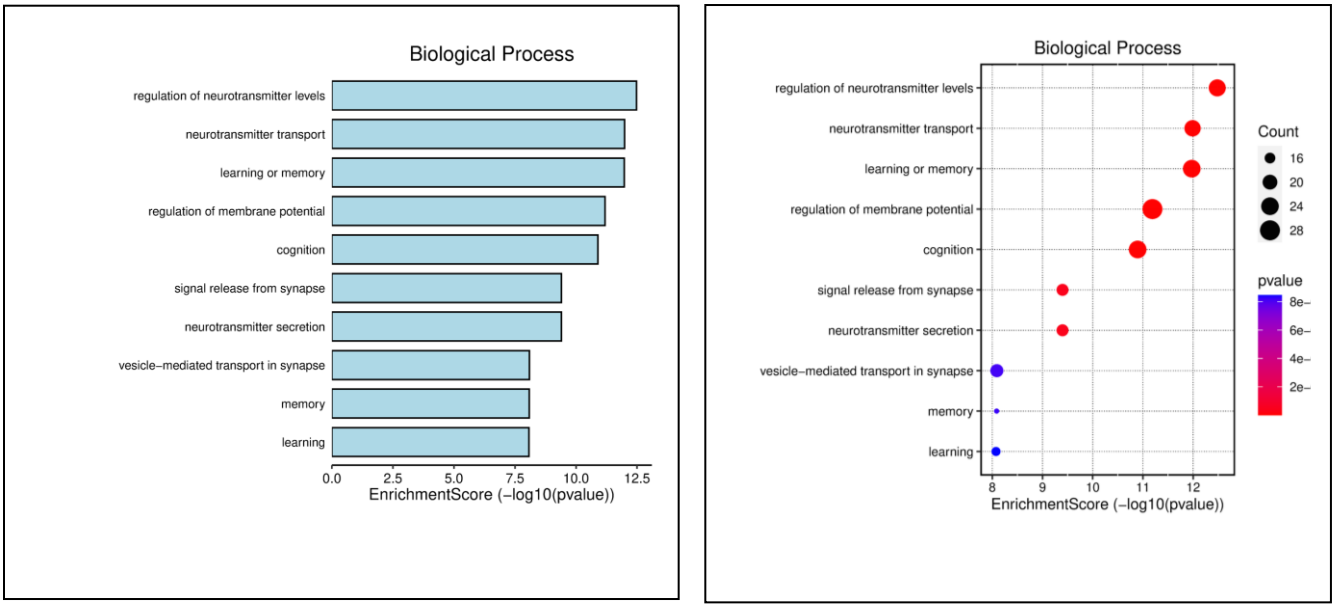


Fig: Enrichment scores of **biological processes** in bar plot and dot plot (+-8)

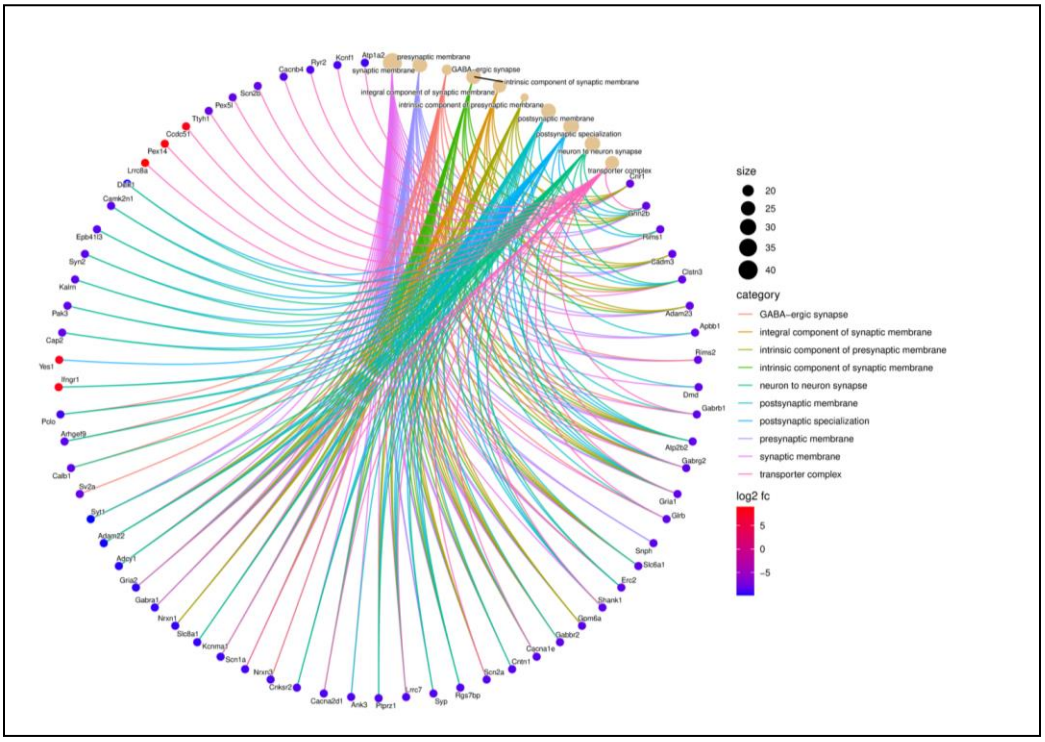


Fig: Network representation of **cellular components** (+-8)

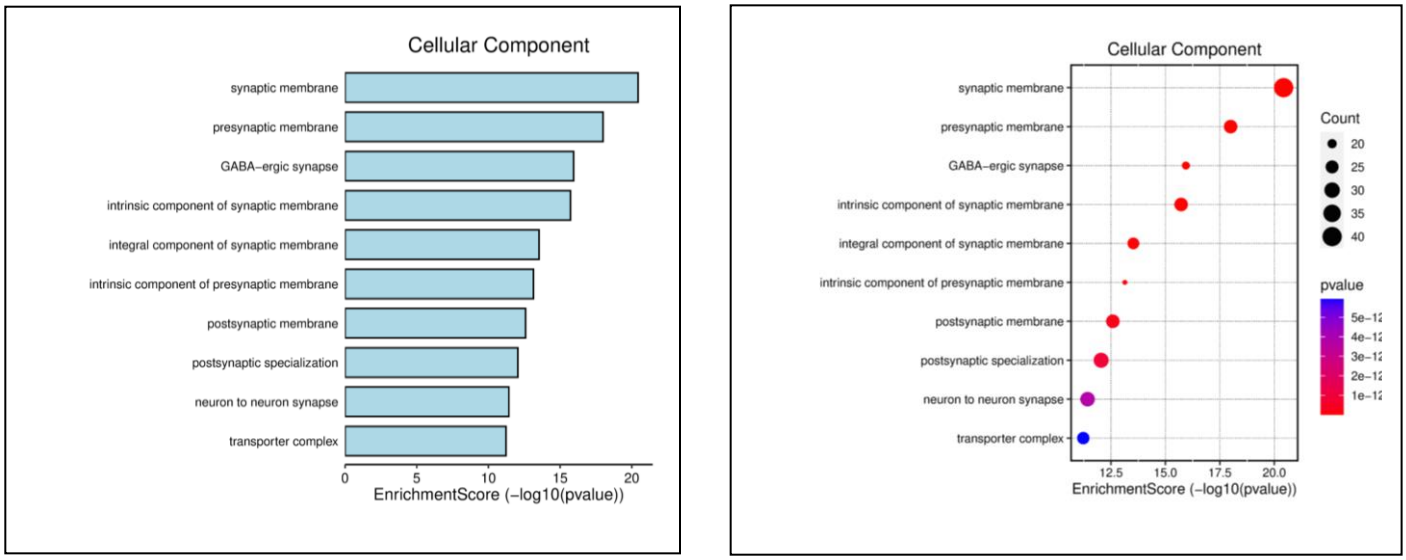


Fig: Enrichment score of **cellular components** in bar plot and dot plot (+-8)

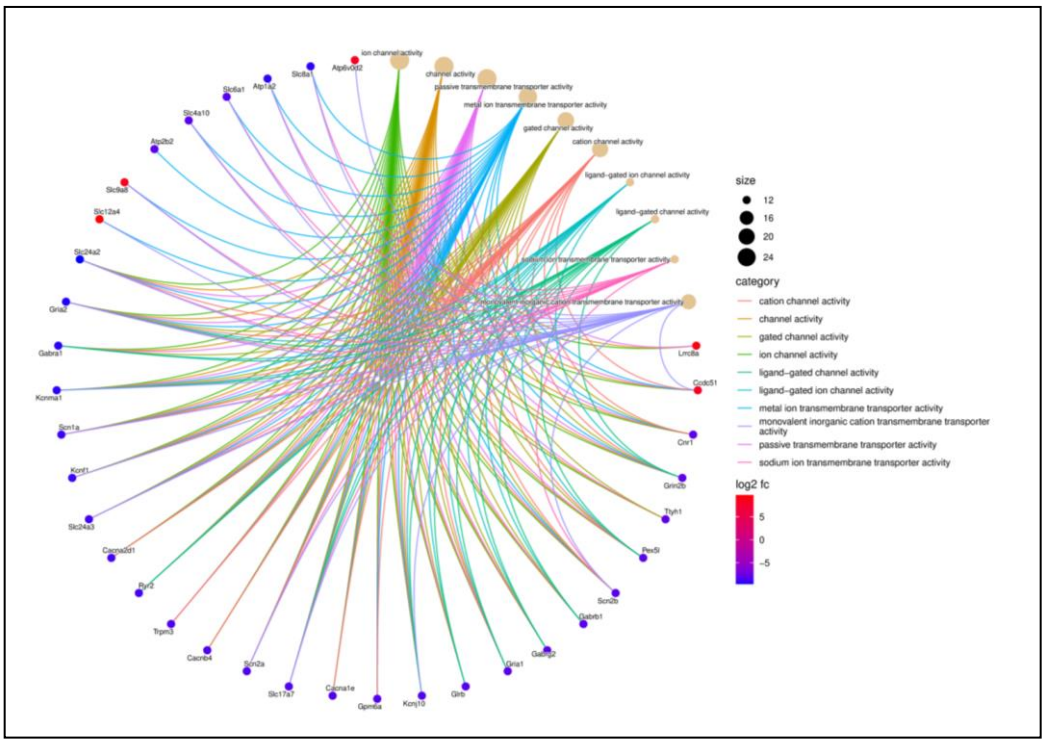
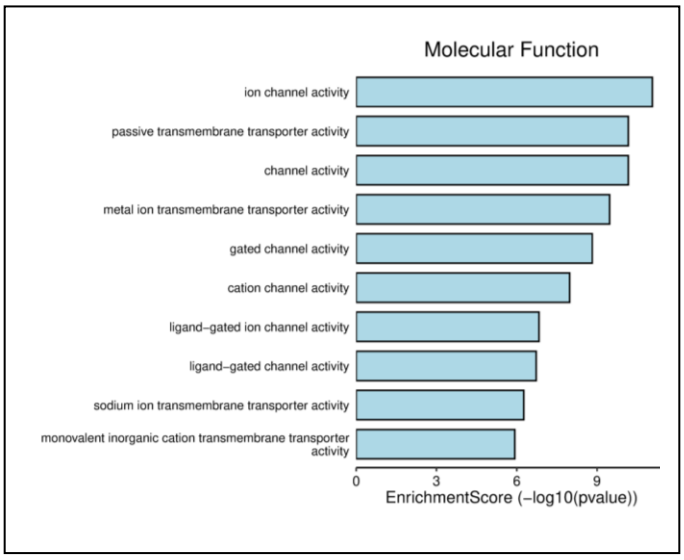


Fig: Network representation of **Molecular functions** (+-8)



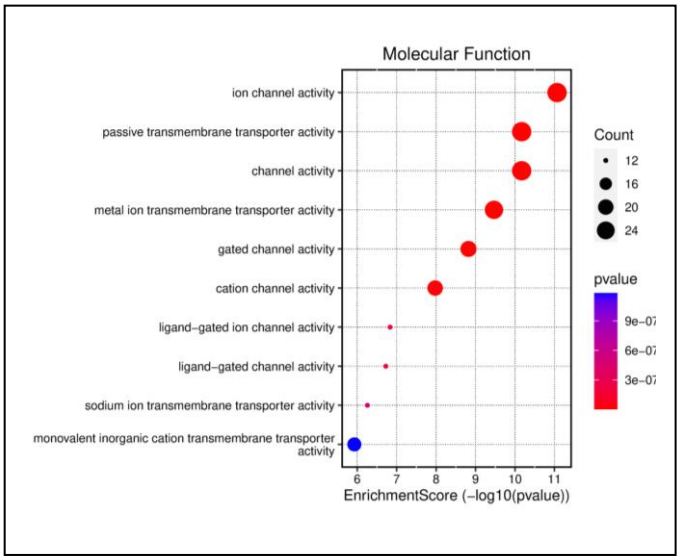


Fig: Enrichment score of **Molecular functions** in bar plot and dot plot (+-8)

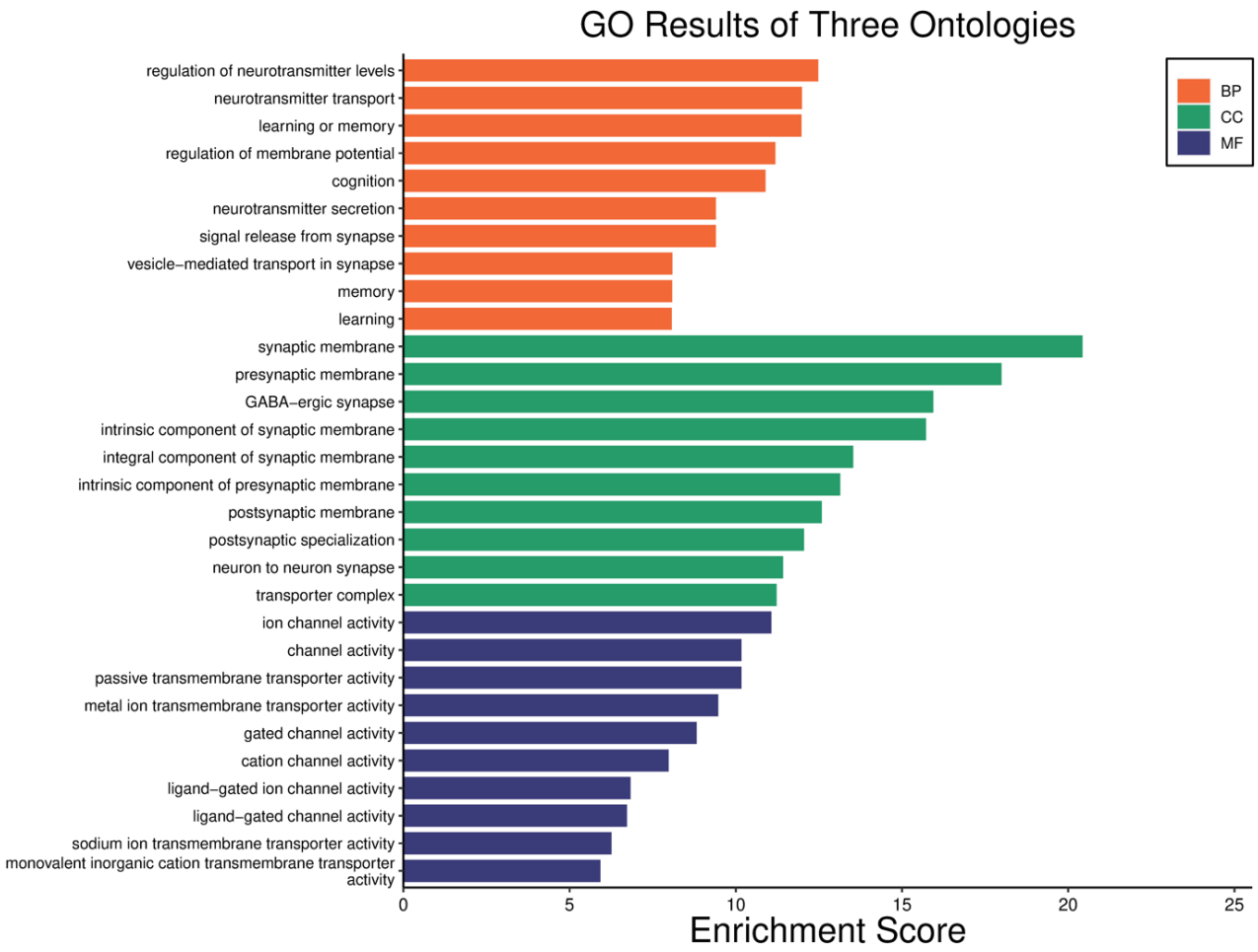


Fig: Enrichment Scores of **three GO ontologies** (+-8)

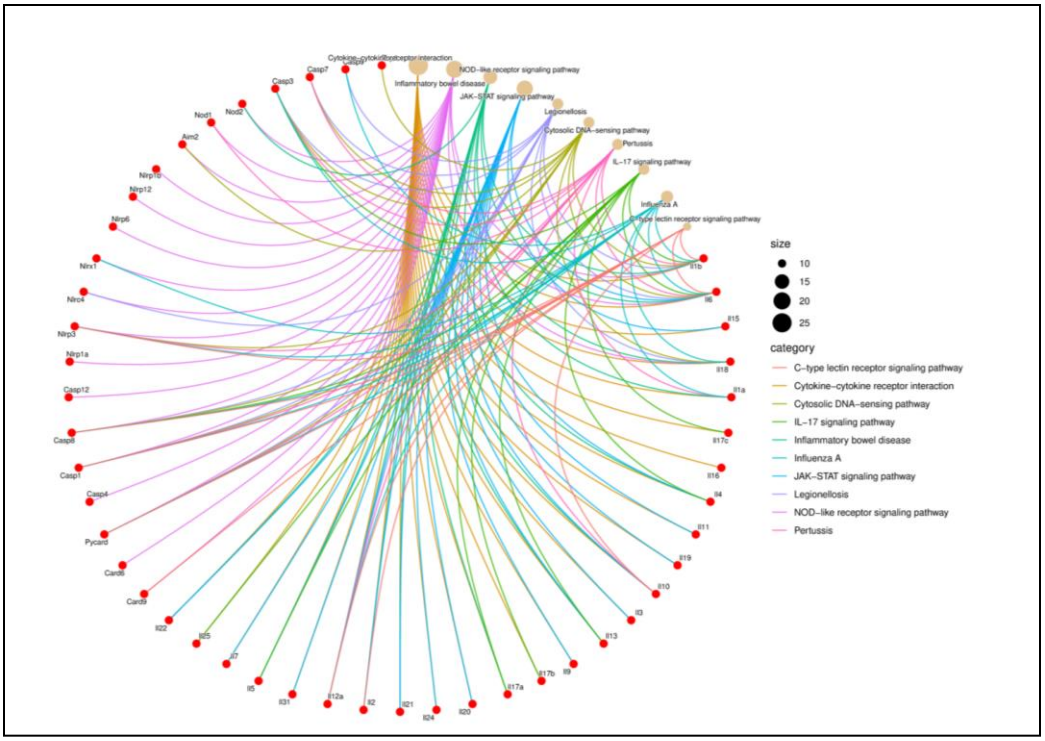
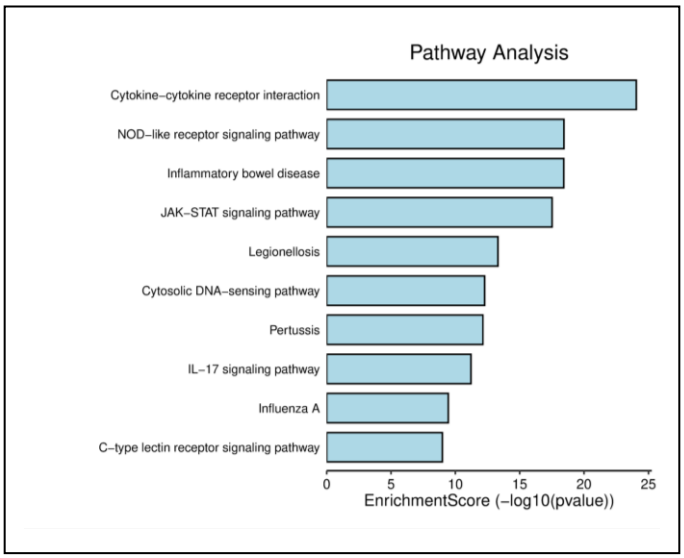


Fig: Network representation of **KEGG pathways for inflammatory genes**



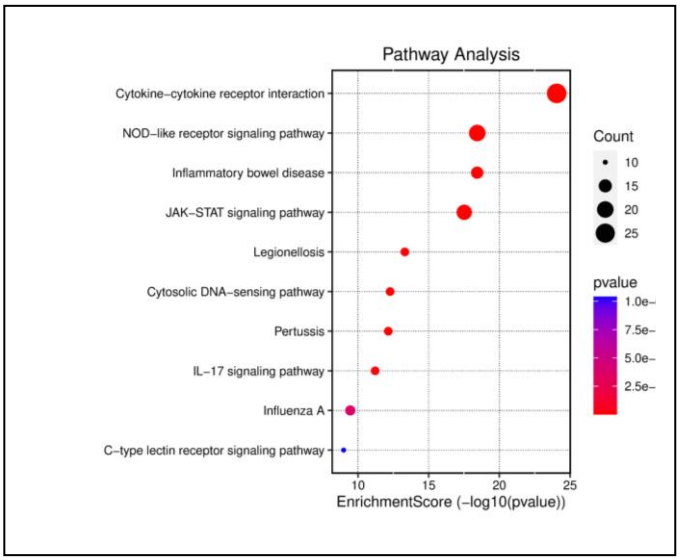


Fig: Enrichment score of **KEGG pathways of Inflammatory genes**

List of all the pathways:

I. **For the set of <= -5 and >= +5**
   1. AGE–RAGE signaling pathway in diabetic complications
   2. Complement and coagulation cascades

II. **For the set of <= -8 and >= +8**
   1. Adrenergic signaling in cardiomyocytes
   2. Amphetamine addiction
   3. cAMP signaling pathway
   4. Circadian entrainment
   5. GABAergic synapse
   6. Insulin secretion
   7. Leishmaniasis
   8. Morphine addiction
   9. Nicotine addiction
   10. Retrograde endocannabinoid signaling

III. **For the set of <= -10 and >= +10**
   1. 2–Oxocarboxylic acid metabolism
   2. Arginine and proline metabolism
   3. Biosynthesis of amino acids
   4. Circadian rhythm
   5. Cysteine and methionine metabolism
   6. Pantothenate and CoA biosynthesis
   7. Protein export
   8. Regulation of lipolysis in adipocytes
   9. Valine, leucine and isoleucine degradation
   10. VEGF signaling pathway

IV. **For the set of inflammation genes**
   1. C–type lectin receptor signaling pathway
   2. Cytokine–cytokine receptor interaction
   3. Cytosolic DNA–sensing pathway
   4. IL–17 signaling pathway
   5. Inflammatory bowel disease
   6. Influenza A
   7. JAK–STAT signaling pathway
   8. Legionellosis
   9. NOD–like receptor signaling pathway
   10. Pertussis

❖ Heat maps

We made heatmaps using **SR plots** by taking up the controls and test samples where for three set of control samples C1, C2, and C3; there was a change only in T1 and T3. The C2 versus T2 sample set has no changes even after the treatment.

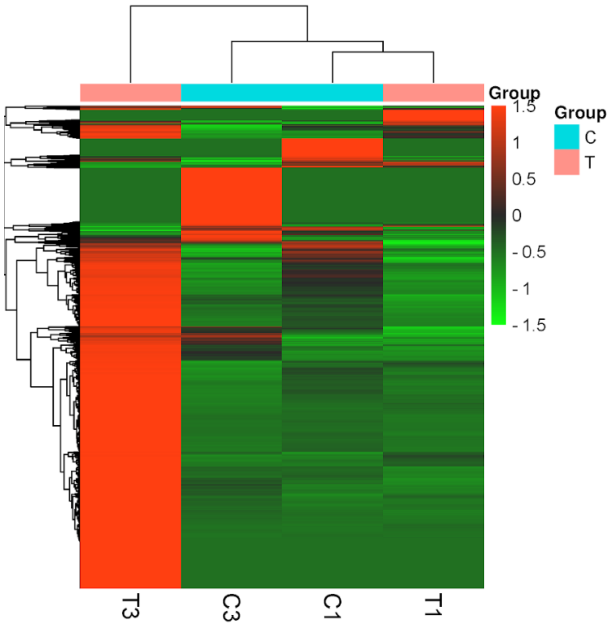


Fig: Heat map for Control versus treats considering only 2 set of samples.

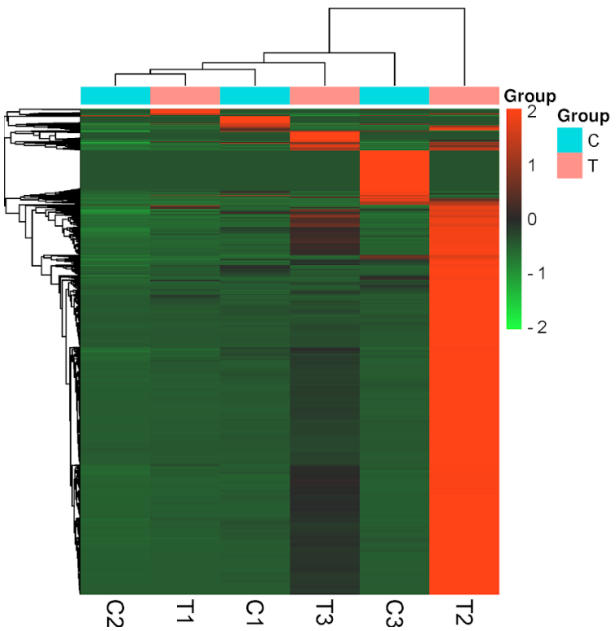


Fig: Heat map for Control versus treats considering all the 3 sample sets.

❖ Gene functional annotations

After performing the differential analysis and data filtration, considering the gene set which are regulated in the range <= -5 and >= +5 we are performing Gene functional annotations, where giving the gene ID as input in the **Genbank, Ensembl, and OMIM softwares**. We are collecting functional role of each gene which includes both coding and non-coding genes.

The drive folder link for the same is:
https://docs.google.com/spreadsheets/d/1-Oo8_dkxZpd3eldwqVXm5lTqoP4LVZVzyoGPTWdiPaU/edit?usp=drive_link

❖ Individual contributions

1. **Rahul naik:** Gene expression counts, Error plots, later moved to work on image data set
2. **Megavath Pavan:** Differential Expression (Volcano plots), Gene expression counts, error plots, data filtration.
3. **Jarpala Ashok:** KEGG pathway analysis, Gene ontology, Enrichment analysis, data filtration, Heatmaps, Gene functional annotations.
4. **Devansh Jatin Shah:** Gene functional annotations
5. **Ritu Badgoti:** Gene functional annotations
6. **Aditya Parkhi:** Gene functional annotations
7. **Pooja Porwal:** Gene functional annotations