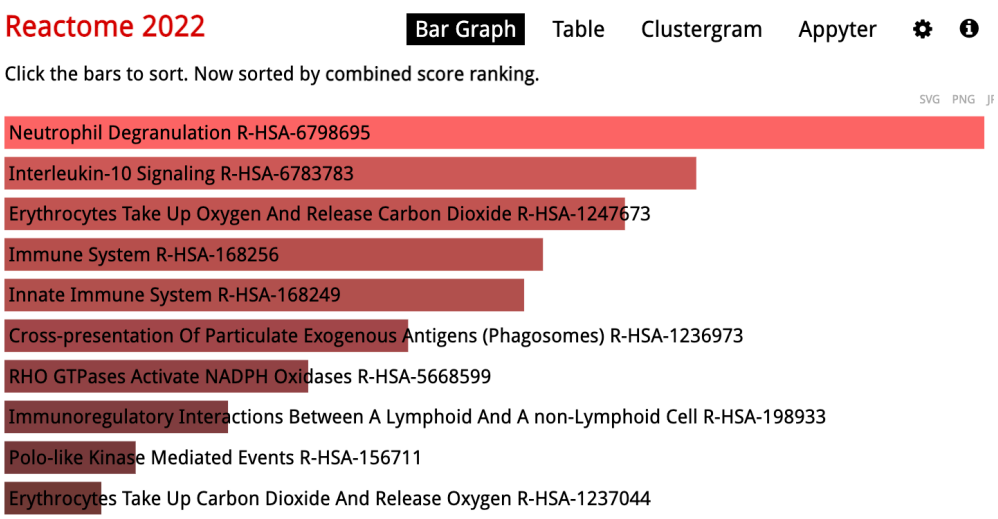
1. As described in the lecture videos, we can derive the up and down tables which correspond to the AML vs Healthy genes (one for genes that are more expressed in patients and the other one for the genes that are less expressed in patients than healthy humans). The code is included in the markdown as well as in the code.R. Also the results are in the files named with AML_vs_Health_Down.txt and AML_vs_Health_Up.txt

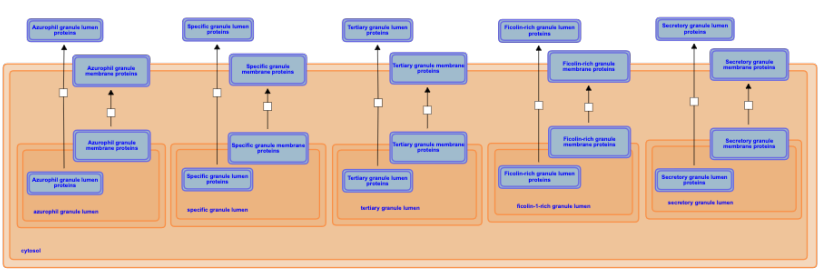
```
#### Differential expression analysis
gr <- factor(gr)
gset$description <- gr
design <- model.matrix(~ description + 0 , gset)
colnames(design) <- levels(gr)
fit <- lmFit(gset, design)
cont.matrix  <- makeContrasts(AML-CD34, levels = design)
fit2 <- contrasts.fit(fit, cont.matrix)
fit2 <- eBayes(fit2, 0.01)
tT <- topTable(fit2, adjust="fdr", sort.by = "B", number = Inf)

tT <- subset(tT, select = c("Gene.symbol",  "Gene.ID", "adj.P.Val", "logFC"))
write.table(tT, "Results/AML_CD34.txt", row.names = F, sep = "\t", quote=F)


aml.up <- subset(tT, logFC > 1 & adj.P.Val < 0.05)
aml.up.genes <- unique(as.character(strsplit2(aml.up$Gene.symbol, "///")))
write.table(aml.up.genes, file="Results/AML_vs_Healthy_UP.txt", quote=F, row.names=F, col.names=F)
aml.down <- subset(tT, logFC < -1 & adj.P.Val < 0.05)
aml.down.genes <- unique(as.character(strsplit2(aml.down$Gene.symbol, "///")))
write.table(aml.down.genes, file="Results/AML_vs_Health_Down.txt", quote=F, row.names=F, col.names=F) # in the output file, we have genes that have
even various names.
```

2. We can start with the Up genes and upload the names in to the enrichr site and derive the following results:

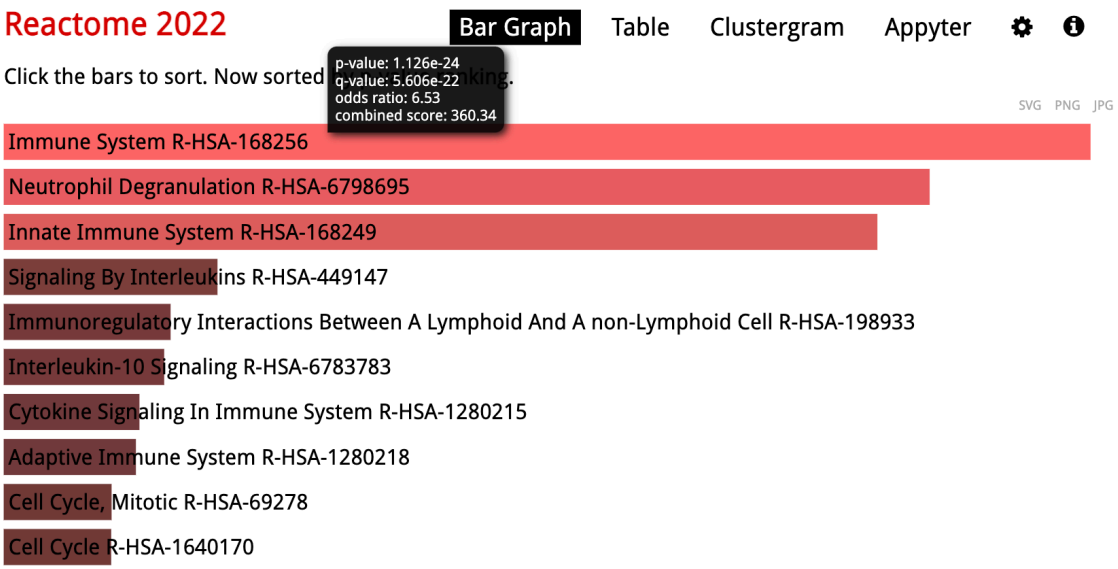
- pathways:
  - Reactome: sorting by the combined score, we can see that the Neutrophil Degranulation, Interleukin-10 rank higher among the other pathways.



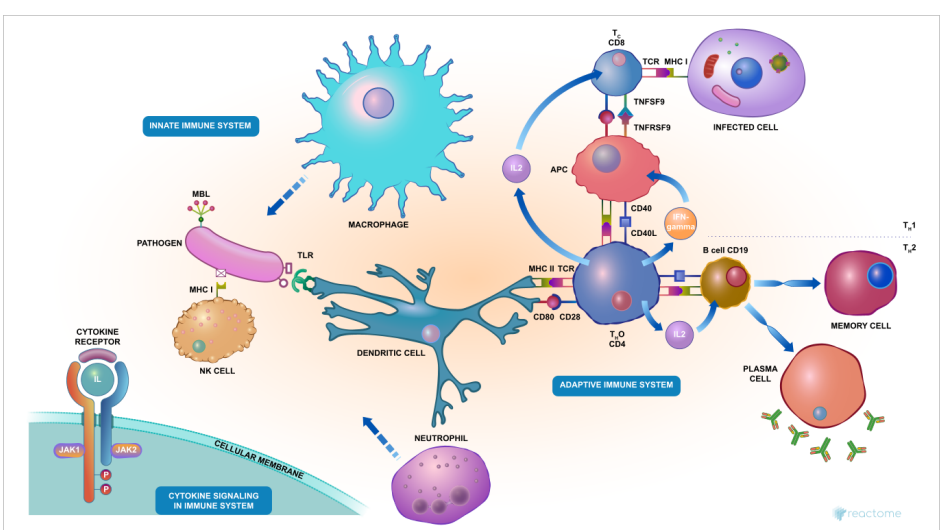
Neutrophils are the most abundant leukocytes (white blood cells), indispensable in defending the body against invading microorganisms. Below is the Neutrophil Degranulation pathway represented in reactome website.



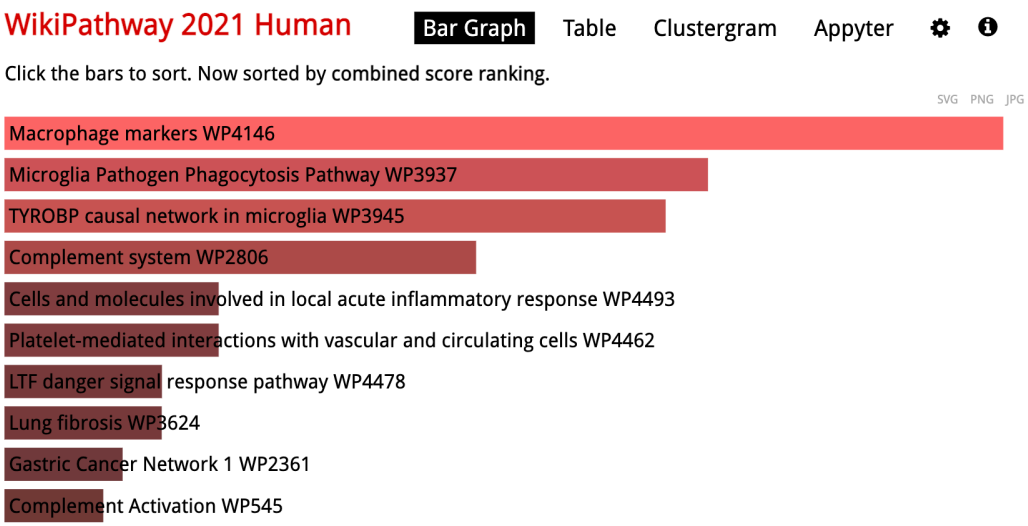
Also sorting by the p-value, we can see that Immune System pathway has much lower value and this makes sense since the genes are for AML a cancer which attacks the immune system.



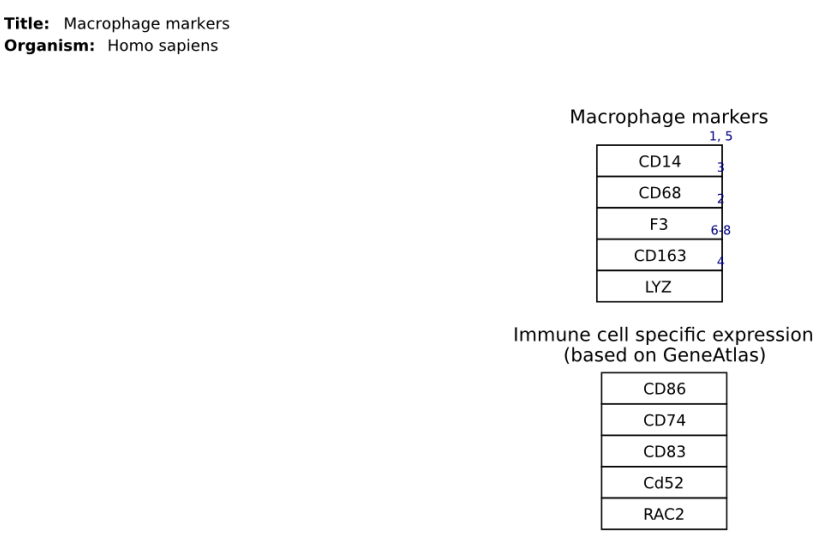
Humans are exposed to millions of potential pathogens daily, through contact, ingestion, and inhalation. Our ability to avoid infection depends on the adaptive immune system and during the first critical hours and days of exposure to a new pathogen, our innate immune system. Immune System path way is shown below.



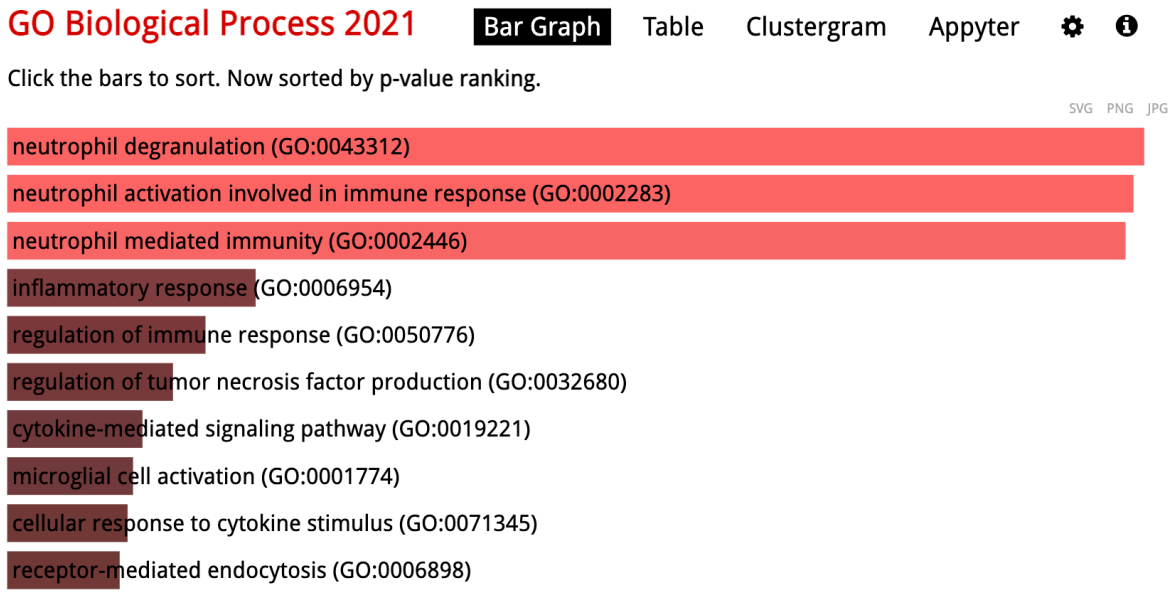
WikiPathways 2021 Human: As shown in the bar graph, Macrophage markers have higher ranking.



Macrophages are cells within the tissues that originate from specific white blood cells called monocytes. Below is the pathway according to the wikipathways data base.



- Ontologies: Gene Ontology (GO) enrichment analysis is ubiquitously used for interpreting high throughput molecular data and generating hypotheses about underlying biological phenomena of experiments. And using the enrichr website, some of the ontologies are as bellow:
    - GO Biological Process 2021: if ranked by p-value, the results are:

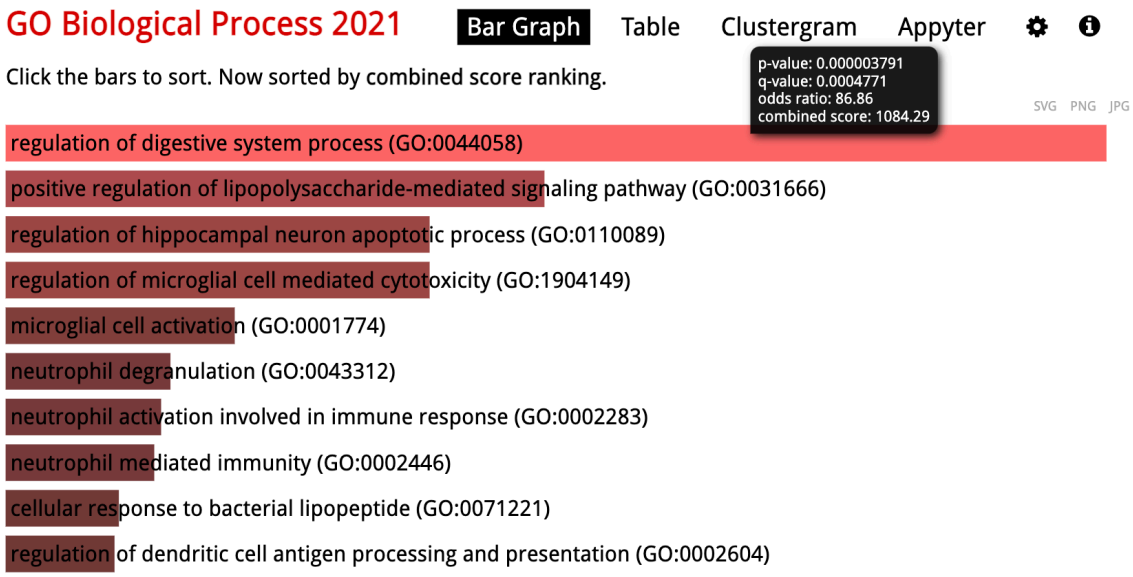


neutrophil degranulation is ranked higher which is defined as:

GO Term Detail

Term:	**neutrophil degranulation**
Synonyms:	heterophil degranulation | neutrophil granule exocytosis
Definition:	The regulated exocytosis of secretory granules containing preformed mediators such as proteases, lipases, and inflammatory mediators by a neutrophil.
Parent Terms:	*is-a* leukocyte degranulation *is-a* neutrophil activation involved in immune response *part-of* neutrophil mediated immunity
Category:	Biological Process
ID:	GO:0043312

Also if we rank the results by the combined score, regulation of digestive system process is ranked higher:

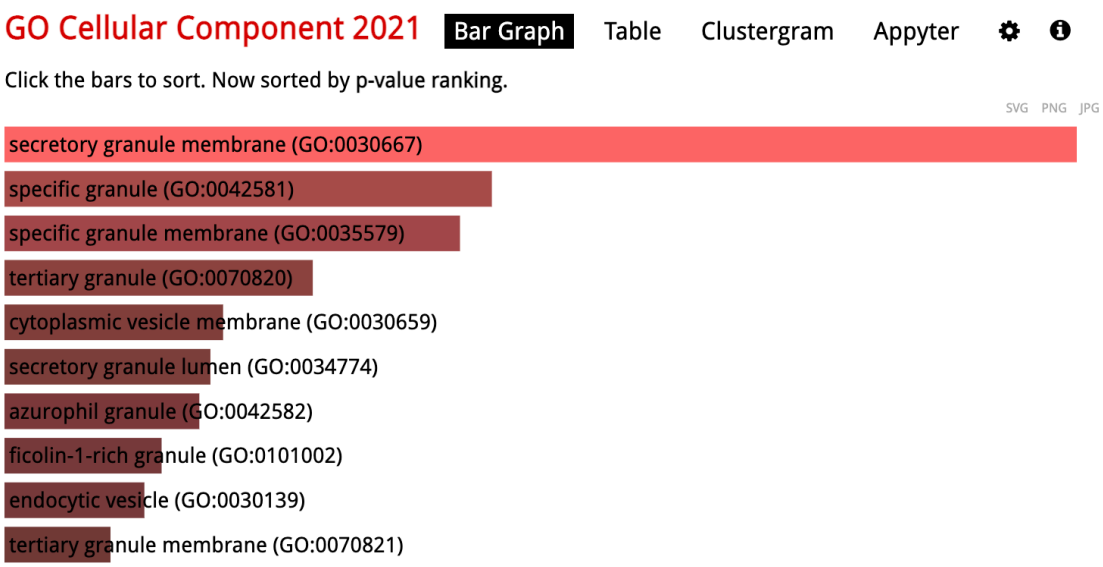


Its definition in the website is as below:

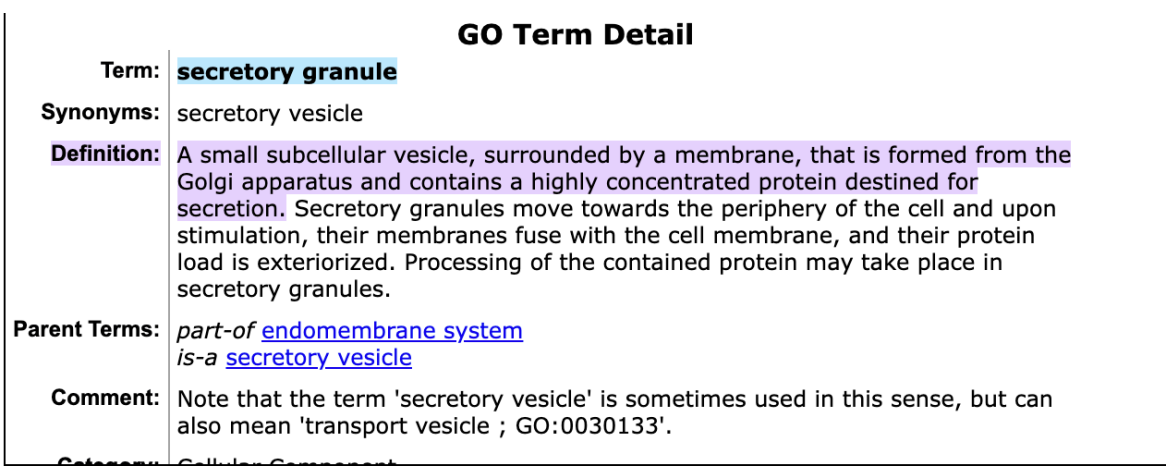
GO Term Detail

Term:	**digestive system process**
Definition:	A physical, chemical, or biochemical process carried out by living organisms to break down ingested nutrients into components that may be easily absorbed and directed into metabolism.
Parent Terms:	*part-of* digestion *is-a* system process
Category:	Biological Process
ID:	GO:0022600

- GO Cellular Process 2021: if ranked by p-value, the results show that secretory granule membrane is ranked higher which is defined as:



Its definition in the website is as below:



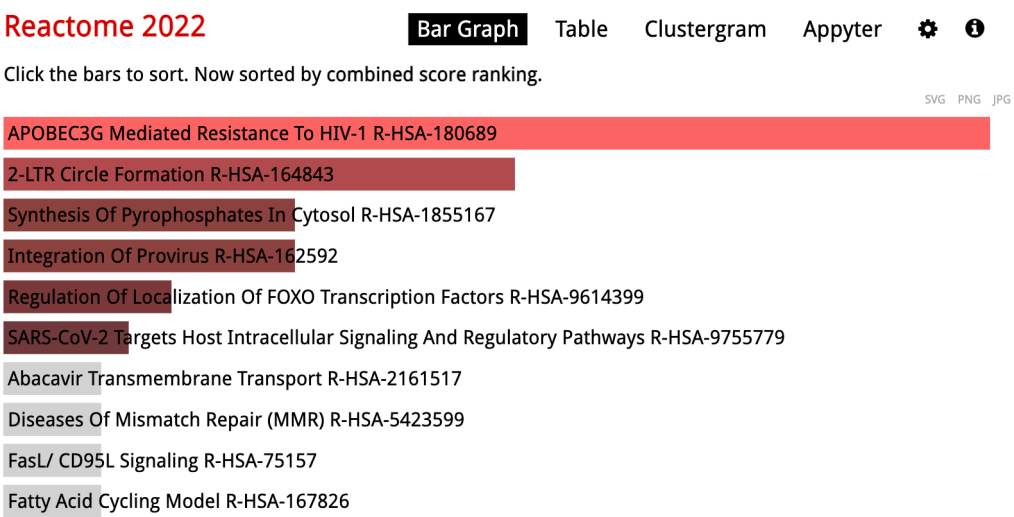
GO Term Detail

Term:	**secretory granule**
Synonyms:	secretory vesicle
Definition:	A small subcellular vesicle, surrounded by a membrane, that is formed from the Golgi apparatus and contains a highly concentrated protein destined for secretion. Secretory granules move towards the periphery of the cell and upon stimulation, their membranes fuse with the cell membrane, and their protein load is exteriorized. Processing of the contained protein may take place in secretory granules.
Parent Terms:	*part-of* endomembrane system *is-a* secretory vesicle
Comment:	Note that the term 'secretory vesicle' is sometimes used in this sense, but can also mean 'transport vesicle ; GO:0030133'.

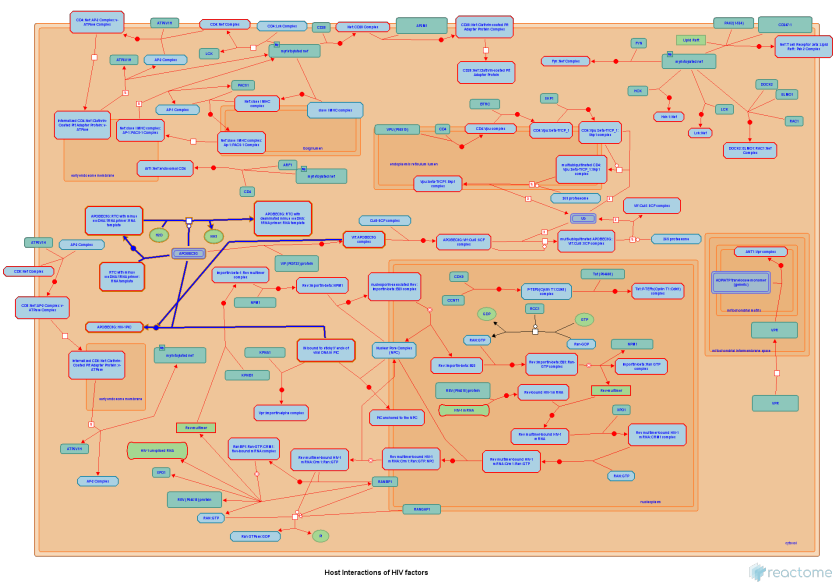
secretory granule membrane is also ranked higher using the combined score.

We can continue with the Down genes and upload the names in to the enrichr site and derive the following results:

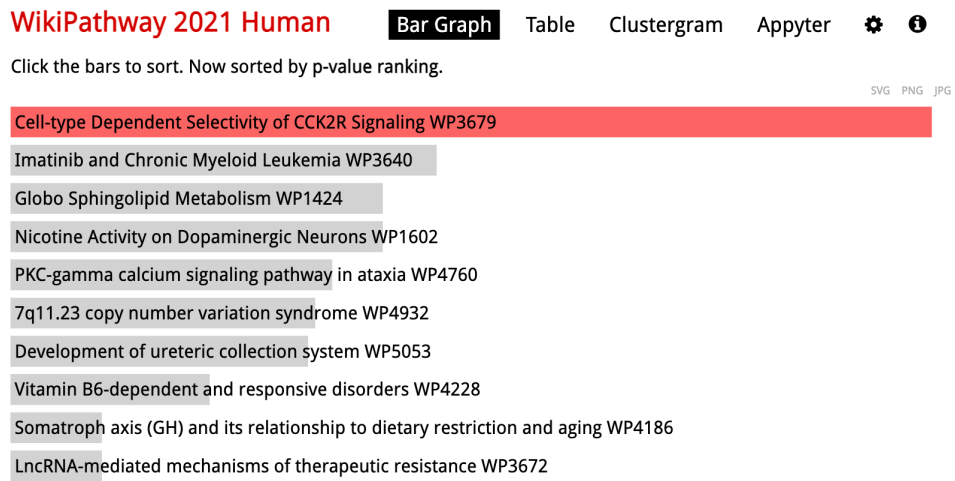
- pathways:
  - Reactome: sorting by the combined score, we can see that the APOBEC3G Mediated Resistance To HIV-1 R-HSA-180689 rank higher among the other pathways. (it's the same if ranked by the p-value)



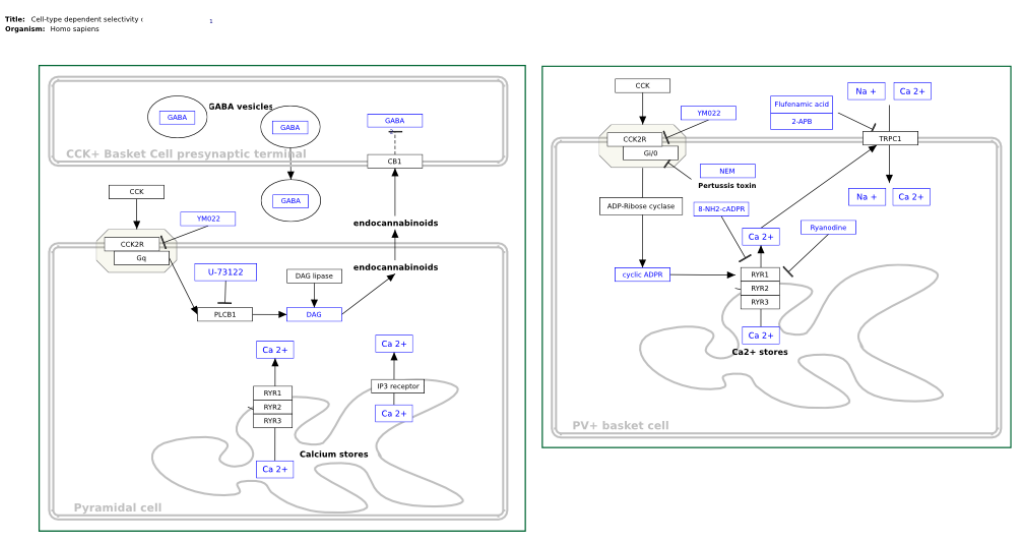
Below is the pathway represented in reactome website.



- WikiPathways 2021 Human: As shown in the bar graph, Cell-type dependent selectivity of CCK2R signaling have higher ranking.

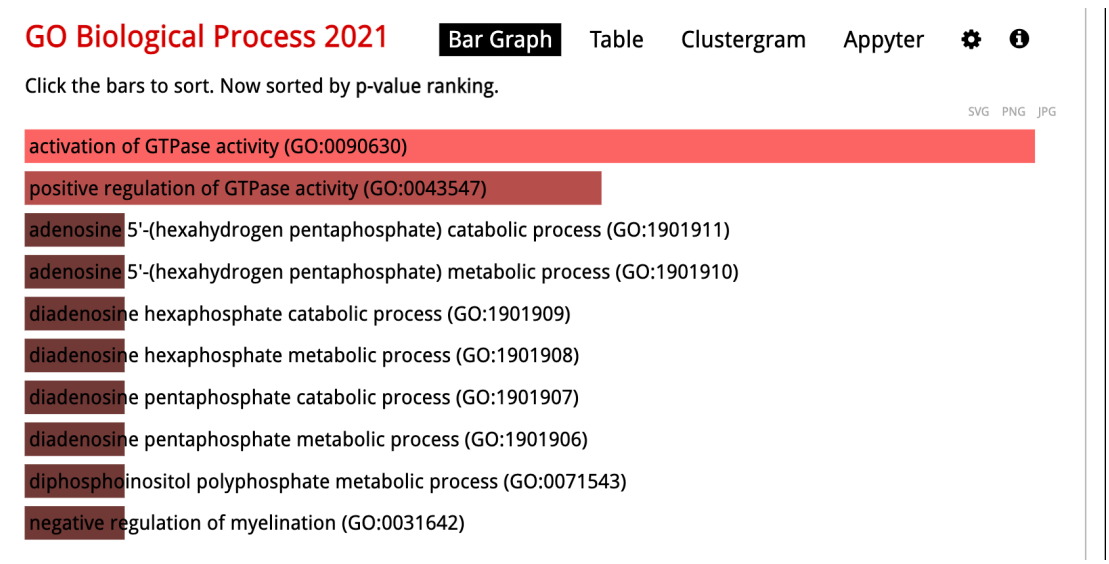


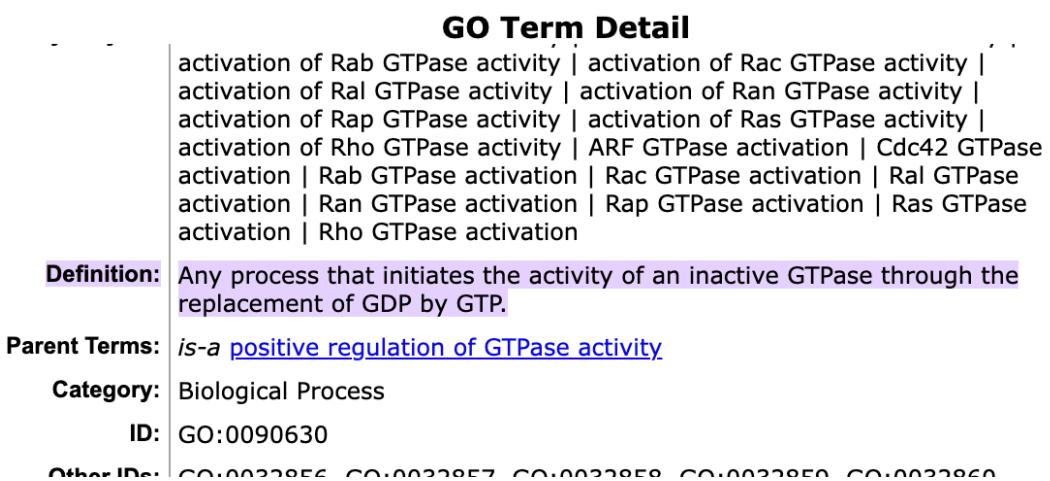
Below is the pathway according to the wikipathways data base.



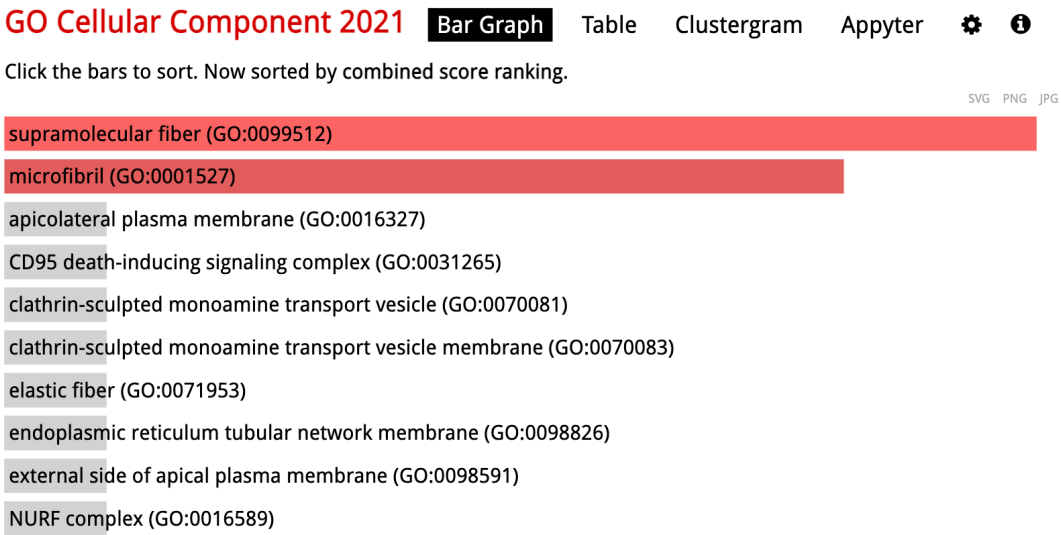
- Ontologies
  - GO Biological Process 2021: if ranked by p-value, the results are:

activation of GTPase activity is ranked higher which is defined as:





- GO Cellular Process 2021: if ranked by p-value, the results show that supramolecular fiber and microfibrill are ranked higher which is defined as:



Its definition in the website is as below:

GO Term Detail

Term:	**supramolecular fiber**
Synonyms:	fibril
Definition:	A polymer consisting of an indefinite number of protein or protein complex subunits that have polymerised to form a fiber-shaped structure.
Parent Terms:	*is-a* supramolecular polymer
Category:	Cellular Component
ID:	GO:0099512
Other IDs:	GO:0043205

3.
a)
for the related **genes**, we can use the articles below:

https://www.ncbi.nlm.nih.gov/pmc/articles/PMC5767295/#:~:text=Some%20of%20the%20identified%20genes,%2C%20ASXL1%2C%20PTPN11%20and%20CBL.

It's been stated that "Recently, with the development of methodologies of massive sequencing, new genetic mutations associated with acute myeloid leukemia have been identified. Some of the identified genes include KIT, FLT3, NPM1, CEBPA, RAS, WT1, BAALC, ERG, MN1, DNMT, TET2, IDH, ASXL1, PTPN11 and CBL."

If we take a look at our UP genes, we can see that many of them are also included above, such as KIT, FLT3, NPM1, WT1 and etc.

https://www.ncbi.nlm.nih.gov/pmc/articles/PMC6142505/

Also according to this article, we have:

- KIT. Nearly exclusively identified in AML with core-binding factor (CBF) translocations, mutations in the receptor tyrosine kinase KIT, particularly the D816V missense variant, occur in ~20% of CBF-AML and are frequently associated with an adverse prognosis...
- FLT3. ITDs and/or activating kinase domain point mutations (D835) in the FLT3 gene are present in nearly one-third of patients with AML, resulting in constitutive activation and downstream signaling through the RAS/RAF/MEK/mammalian target of rapamycin growth and proliferation pathways, and the phosphatidylinositol 3 kinase (PI3K)/AKT prosurvival pathway...

And both exist in the UP genes list which verifies our findings.

https://pubmed.ncbi.nlm.nih.gov/32203165/

Moreover, according to this source, CD83 which is also present in our file, CD82 drives acute myeloid leukemia chemoresistance by modulating protein kinase C alpha and β1 integrin activation.

"Analysis of the TARGET and BEAT AML databases identifies a correlation between CD82 expression and overall survival of AML patients."

Also for the found ontologies found in the previous section we have:

- **pathways**

Neutrophil Degranulation

https://www.uptodate.com/contents/acute-myeloid-leukemia-aml-treatment-in-adults-beyond-the-basics/print

The overgrowth of these cells leads to an inadequate number of normal, healthy blood cells, including white blood cells, red blood cells, and platelets. This can result in: Neutropenia (low numbers of neutrophils) – Neutrophils are a type of white blood cell that helps to fight infection.

immune system
AML cells evade or suppress the immune system through five main mechanisms: reduced expression of major histocompatibility complex (MHC) molecules, enhanced inhibitory ligand expression, reduced activating ligand/receptor expression, ligand shedding, and manipulation of soluble factors within the microenvironment

- **gene ontology**

Macrophage
AML cells induce the expansion and/or migration of tissue-resident macrophages. They function as AAMs since they support the growth of AML cells both in vivo and in vitro. Furthermore, the polarization of AAMs depends on the presence of Gfi1, which is a potential new regulator of AAMs and macrophage polarization
https://www.ncbi.nlm.nih.gov/pmc/articles/PMC5046651/#:~:text=AML%20cells%20induce%20the%20expansion,of%20AAMs%20and%20macrophage%20polarization.

b)We found FLT3 in our gene tables that this gene helps the cells make a protein (also called FLT3) which helps the cells grow. and for some patients whose leukemia cells have an FLT3 gene mutation, the targeted therapy. drug midostaurin (Rydapt) might be given along with chemo.

Drugs that target the FLT3 protein can help treat some of these leukemias.
Midostaurin (Rydapt) is a drug that works by blocking FLT3 and several other proteins on cancer cells that can help the cells grow. This drug can be used along with certain chemotherapy drugs to treat newly diagnosed adults whose leukemia cells have a mutation in the FLT3 gene.

https://www.cancer.org/cancer/acute-myeloid-leukemia/treating/typical-treatment-of-aml.html

In the United States, gemtuzumab ozogamicin is indicated for newly diagnosed CD33-positive acute myeloid leukemia (AML) for adults and children one month and older and for the treatment of relapsed or refractory CD33-positive AML in adults and children two years and older

We found Bcl-2 in our gene tables that Bcl-2 (B-cell lymphoma 2), encoded in humans by the BCL2 gene, is the founding member of the Bcl-2 family of regulator proteins that regulate cell death (apoptosis), by either inhibiting (anti-apoptotic) or inducing (pro-apoptotic) apoptosis. It was the first apoptosis regulator identified in any organism. Bcl-2 derives its name from B-cell lymphoma 2, as it is the second member of a range of proteins initially described in chromosomal translocations involving chromosomes 14 and 18 in follicular lymphomas. Orthologs (such as Bcl2 in mice) have been identified in numerous mammals for which complete genome data are available.
Venclexta is a type of targeted therapy for AML. It works by targeting and blocking the action of a specific protein within leukemia cells and decreases its amounts in cell.

Bahar OverisGharan 98106242
Zahra Soukhtehdel 98105138
Lachin Naghashyar 98110179