

Papalex CUL3 Analysis

2023-02-21

Goal

The aim of this report is to replicate the results of Papalex et al's pathway enrichment analysis.

```
# Load packages.
library(Seurat)

## Attaching SeuratObject

library(SeuratData)

## -- Installed datasets ----- SeuratData v0.2.2 --
## v thp1.eccite 3.1.5

## ----- Key -----

## v Dataset loaded successfully
## > Dataset built with a newer version of Seurat than installed
## (?) Unknown version of Seurat installed

library(ggplot2)
library(patchwork)
library(scales)
library(dplyr)

## 
## Attaching package: 'dplyr'

## The following objects are masked from 'package:stats':
## 
##     filter, lag

## The following objects are masked from 'package:base':
## 
##     intersect, setdiff, setequal, union

library(reshape2)
library(mixtools)

## mixtools package, version 2.0.0, Released 2022-12-04
## This package is based upon work supported by the National Science Foundation under Grant No. SES-051
```

```

library(stringr)
library(enrichR)

## Welcome to enrichR
## Checking connection ...

## Enrichr ... Connection is Live!
## FlyEnrichr ... Connection is available!
## WormEnrichr ... Connection is available!
## YeastEnrichr ... Connection is available!
## FishEnrichr ... Connection is available!
## OxEnrichr ... Connection is available!

library(kableExtra)

```

```

##
## Attaching package: 'kableExtra'
##
## The following object is masked from 'package:dplyr':
##
##     group_rows

```

```
library(varhandle)
```

Download Data

```

# Download dataset using SeuratData.
options(timeout = 1000)
InstallData(ds = "thp1.eccite")

## Warning: The following packages are already installed and will not be
## reinstalled: thp1.eccite

# Setup custom theme for plotting.
custom_theme <- theme(
  plot.title = element_text(size=16, hjust = 0.5),
  legend.key.size = unit(0.7, "cm"),
  legend.text = element_text(size = 14))

# Load object.
eccite <- LoadData(ds = "thp1.eccite")

```

Normalize and plot UMAP

```

# Normalize protein.
eccite <- NormalizeData(
  object = eccite,
  assay = "ADT",
  normalization.method = "CLR",
  margin = 2)

## Normalizing across cells

# Prepare RNA assay for dimensionality reduction:
# Normalize data, find variable features and scale data.
DefaultAssay(object = eccite) <- 'RNA'
eccite <- NormalizeData(object = eccite) %>% FindVariableFeatures() %>% ScaleData()

## Centering and scaling data matrix

# Run Principle Component Analysis (PCA) to reduce the dimensionality of the data.
eccite <- RunPCA(object = eccite)

## PC_ 1
## Positive: BIRC5, TOP2A, CDC20, MKI67, CENPF, TPX2, CDKN3, UBE2C, CKS1B, NUF2
##          CCNA2, NUSAP1, KIAA0101, CENPA, HMGB2, SGOL1, TYMS, STMN1, MYBL2, GTSE1
##          ASPM, H2AFZ, CDCA2, HMMR, CDCA8, KIF2C, CKAP2L, PTTG1, MND1, UBE2T
## Negative: FTH1, FCER1G, NEAT1, SOD2, FTL, MAFB, BTG1, NPC2, CTSL, CTSC
##          CTSB, SLC31A2, CHI3L1, FAM26F, TNFSF13B, GBP5, PLAUR, EVL, GK, ASAHI
##          HLA-DRB1, HLA-DRA, SPP1, SCPEP1, CD74, SAT1, GBP1, SLAMF7, WARS, SDS
## PC_ 2
## Positive: HYOU1, PDIA4, HSPA5, SDF2L1, MEI1, MANF, DNAJB9, NUCB2, TRIB3, WIPI1
##          CRELD2, HSP90B1, MSTO1, SLC39A14, HERPUD1, ALDH1L2, DERL3, VIMP, SEC11C, SERP1
##          PPAPDC1B, CDK2AP2, OSTC, DNAJB11, ERO1LB, SEC61G, SYVN1, TMED2, DNAJC3, PYCR1
## Negative: HSPA8, KIAA0101, TYMS, MKI67, FCER1G, CHI3L1, ACTG1, TOP2A, MYBL2, HSP90AA1
##          CCNA2, BIRC5, CLSPN, PKMYT1, NPC2, NUSAP1, HMGN2, ZWINT, CENPF, H2AFZ
##          TMEM106C, CENPW, TUBA1B, STMN1, CTSC, ASF1B, CDCA5, HMGA1, RRM2, GTSE1
## PC_ 3
## Positive: CDKN1A, ATF5, WARS, PLEK, CXCL10, IL1RN, SOD2, FAM26F, SLC31A2, GBP1
##          IDO1, SLAMF7, GK, HLA-DRA, ISG20, ICAM1, CD274, CCL2, ATF3, GBP5
##          CCL8, CD74, MTHFD2, IL8, FCER1G, GCH1, TNFSF13B, IL4I1, GLUL, RALA
## Negative: QPRT, S100A4, RPLPO, S100A6, ZFP36L2, ALOX5AP, SORL1, ANTXR1, C1orf162, VCAN
##          GLIPR1, CD1D, ID1, CAPN2, ID2, TGFBR1, RGS16, TKT, ITM2C, CDKN2C
##          HSPB1, ACTG1, CORO1A, SMYD3, ID3, RPSA, ALDH2, FOS, AZU1, THYN1
## PC_ 4
## Positive: RMDN3, GCHFR, GRN, DNASE2, WARS, SCCPDH, PSME2, LIPG, CTSD, HLA-DRB1
##          C19orf59, TSPO, HLA-DRB5, LTA4H, HLA-A, IFI30, AGT, GBP5, CEBPE, APOC1
##          GLUL, MARC1, CD74, CD1D, PPARG, ALOX5AP, CLDN23, CD68, S100A8, PLIN2
## Negative: CCL2, IGFBP3, PEA15, CCL3, NFKBIA, MMP9, CCL4, CCL5, POU2F2, IL1B
##          MARCKSL1, CXCL11, MX2, RGS1, CXCL9, USP18, PDPN, SPP1, CLEC5A, E2F1
##          TGFBR1, CKB, RUNX3, PTPN14, SMYD3, TGFBI, CCL8, TESC, GINS2, PNRC1
## PC_ 5
## Positive: FTL, FABP5, PLIN2, CSTB, CTSD, CD36, FTH1, HMGA1, SPOCD1, RMDN3
##          APOC1, DDIT4L, AGPAT9, CDK4, MSR1, E2F1, SRM, GCHFR, GINS2, CLDN23
##          TOMM40, RND3, NCF2, FAM111B, DTL, APOE, STRA13, IL8, SLC11A1, CHCHD10

```

```

## Negative: TNFSF10, NCF1, PSMB9, CXCL11, CXCL10, IFI27, TMEM176B, RARRES3, TMEM176A, NFKBIA
## IL32, SOCS1, IFITM1, GBP1, CCNB1, PLK1, HMMR, PNRC1, PSME2, RGS16
## MYO1G, CD74, ISG20, TNFSF13B, CXCL9, TMEM50B, CDC20, IFIT2, PTTG1, GLIPR1

# Run Uniform Manifold Approximation and Projection (UMAP) to visualize clustering in 2-D.
eccite <- RunUMAP(object = eccite, dims = 1:40)

## Warning: The default method for RunUMAP has changed from calling Python UMAP via reticulate to the R
## To use Python UMAP via reticulate, set umap.method to 'umap-learn' and metric to 'correlation'
## This message will be shown once per session

## 10:14:36 UMAP embedding parameters a = 0.9922 b = 1.112

## 10:14:36 Read 20729 rows and found 40 numeric columns

## 10:14:36 Using Annoy for neighbor search, n_neighbors = 30

## 10:14:36 Building Annoy index with metric = cosine, n_trees = 50

## 0%   10    20    30    40    50    60    70    80    90   100%
## [----|----|----|----|----|----|----|----|----|----|----|-----]

## *****
## 10:14:37 Writing NN index file to temp file /var/folders/tf/8hspl416b70psyc3zvyszfxm0000gn/T//RtmpFW
## 10:14:37 Searching Annoy index using 1 thread, search_k = 3000
## 10:14:41 Annoy recall = 100%
## 10:14:41 Commencing smooth kNN distance calibration using 1 thread with target n_neighbors = 30
## 10:14:41 Initializing from normalized Laplacian + noise (using irlba)
## 10:14:42 Commencing optimization for 200 epochs, with 910832 positive edges
## 10:14:50 Optimization finished

# Generate plots to check if clustering is driven by biological replicate ID,
# cell cycle phase or target gene class.
p1 <- DimPlot(
  object = eccite,
  group.by = 'replicate',
  label = F,
  pt.size = 0.2,
  reduction = "umap", cols = "Dark2", repel = T) +
  scale_color_brewer(palette = "Dark2") +
  ggtitle("Biological Replicate") +
  xlab("UMAP 1") +
  ylab("UMAP 2") +
  custom_theme

## Scale for colour is already present.
## Adding another scale for colour, which will replace the existing scale.

```

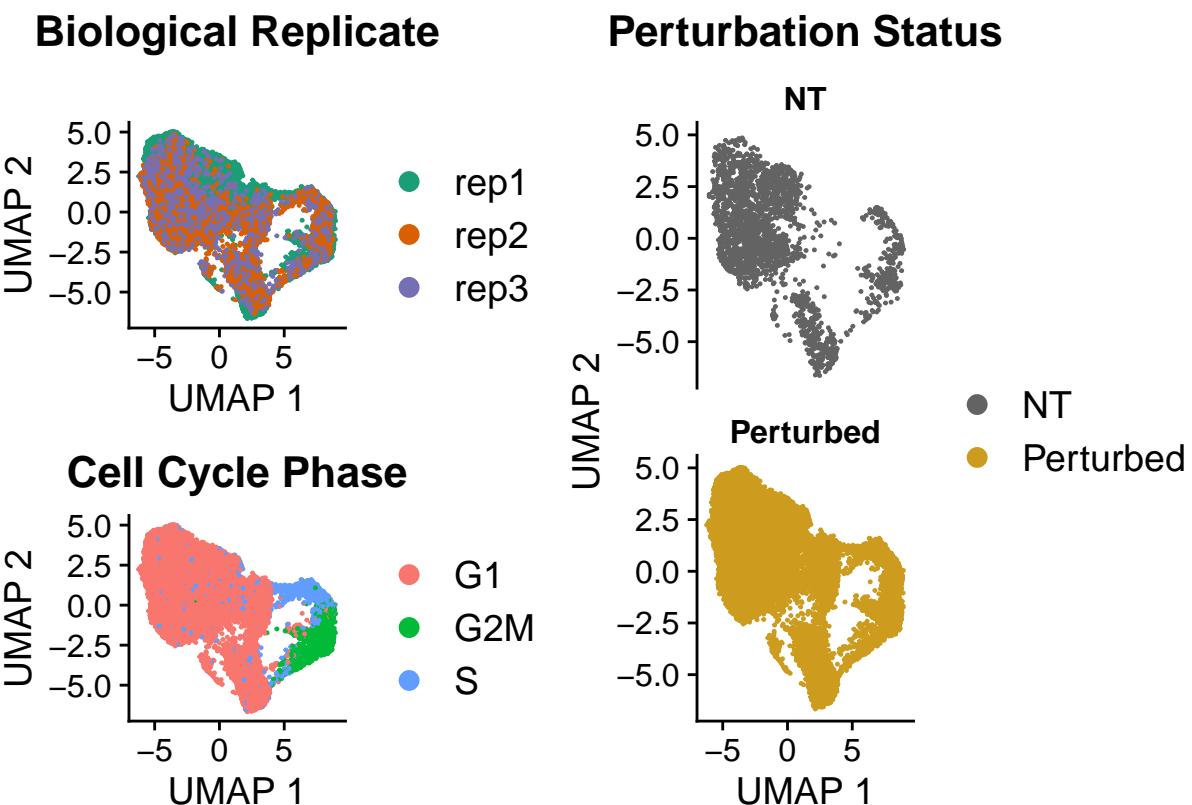
```

p2 <- DimPlot(
  object = eccite,
  group.by = 'Phase',
  label = F, pt.size = 0.2,
  reduction = "umap", repel = T) +
  ggtitle("Cell Cycle Phase") +
  ylab("UMAP 2") +
  xlab("UMAP 1") +
  custom_theme

p3 <- DimPlot(
  object = eccite,
  group.by = 'crispr',
  pt.size = 0.2,
  reduction = "umap",
  split.by = "crispr",
  ncol = 1,
  cols = c("grey39","goldenrod3")) +
  ggtitle("Perturbation Status") +
  ylab("UMAP 2") +
  xlab("UMAP 1") +
  custom_theme

# Visualize plots.
((p1 / p2 + plot_layout(guides = 'auto')) | p3 )

```



Remove technical factors

```
# Calculate perturbation signature (PRTB).
eccite<- CalcPerturbSig(
  object = eccite,
  assay = "RNA",
  slot = "data",
  gd.class ="gene",
  nt.cell.class = "NT",
  reduction = "pca",
  ndims = 40,
  num.neighbors = 20,
  split.by = "replicate",
  new.assay.name = "PRTB")

## Processing rep1

## Processing rep3

## Processing rep2

# Prepare PRTB assay for dimensionality reduction:
# Normalize data, find variable features and center data.
DefaultAssay(object = eccite) <- 'PRTB'

# Use variable features from RNA assay.
VariableFeatures(object = eccite) <- VariableFeatures(object = eccite[["RNA"]])
eccite <- ScaleData(object = eccite, do.scale = F, do.center = T)

## Centering data matrix

# Run PCA to reduce the dimensionality of the data.
eccite <- RunPCA(object = eccite, reduction.key = 'prtbpca', reduction.name = 'prtbpca')

## Warning: Keys should be one or more alphanumeric characters followed by an
## underscore, setting key from prtbpca to prtbpca_

## Warning: All keys should be one or more alphanumeric characters followed by an
## underscore '_', setting key to prtbpca_

## prtbpca_ 1
## Positive: SPP1, S100A4, RPLPO, VCAN, ZFP36L1, TREM2, TGFBR1, CAPN2, TGFB1, LGALS1
##           RPSA, SORL1, FSCN1, CSF1R, YWHAH, LMNA, RPS2, ADORA3, HSPB1, CORO1A
##           ID2, MMP9, VAT1, GLO1, COL6A1, AP1S2, NFKBIA, MGST3, APOE, IL8
## Negative: CD74, HLA-DRA, CXCL10, WARS, GBP5, GBP1, IFI27, HLA-DRB1, FAM26F, PSMB9
##            HLA-DRB5, IL18BP, PSME2, SOCS1, HLA-DPA1, HLA-DQB1, SOD2, IFITM1, NCF1, S100A8
##            HLA-A, GLUL, CTSL, CD70, FCGR1B, HLA-DMA, HLA-DPB1, FCER1G, LY6E, CHI3L1
##            prtbpca_ 2
## Positive: CXCL10, CXCL11, CXCL9, GBP1, SOCS1, GBP5, SOD2, TNFSF13B, CCL2, IFIT3
##            IL32, MX1, GYPC, IL18BP, ISG20, WARS, TNFSF10, IDO1, LY6E, IFI27
```



```

## Warning: Keys should be one or more alphanumeric characters followed by an
## underscore, setting key from prtbumap to prtbumap_

## Warning: All keys should be one or more alphanumeric characters followed by an
## underscore '_', setting key to prtbumap_

# Generate plots to check if clustering is driven by biological replicate ID,
# cell cycle phase or target gene class.
q1 <- DimPlot(
  object = eccite,
  group.by = 'replicate',
  reduction = 'prtbumap',
  pt.size = 0.2, cols = "Dark2", label = F, repel = T) +
  scale_color_brewer(palette = "Dark2") +
  ggtitle("Biological Replicate") +
  ylab("UMAP 2") +
  xlab("UMAP 1") +
  custom_theme

## Scale for colour is already present.
## Adding another scale for colour, which will replace the existing scale.

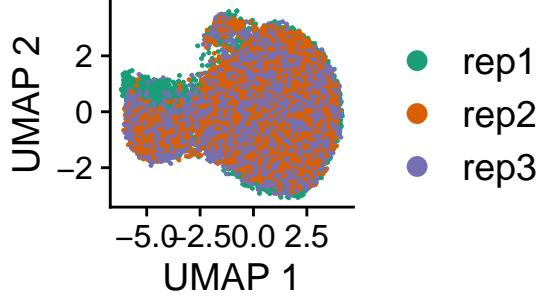
q2 <- DimPlot(
  object = eccite,
  group.by = 'Phase',
  reduction = 'prtbumap',
  pt.size = 0.2, label = F, repel = T) +
  ggtitle("Cell Cycle Phase") +
  ylab("UMAP 2") +
  xlab("UMAP 1") +
  custom_theme

q3 <- DimPlot(
  object = eccite,
  group.by = 'crispr',
  reduction = 'prtbumap',
  split.by = "crispr",
  ncol = 1,
  pt.size = 0.2,
  cols = c("grey39","goldenrod3")) +
  ggtitle("Perturbation Status") +
  ylab("UMAP 2") +
  xlab("UMAP 1") +
  custom_theme

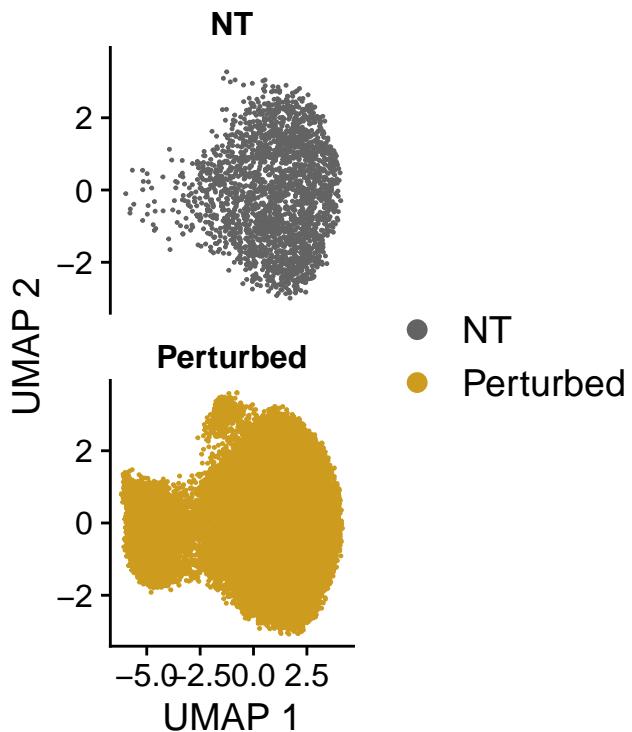
# Visualize plots.
(q1 / q2 + plot_layout(guides = 'auto') | q3)

```

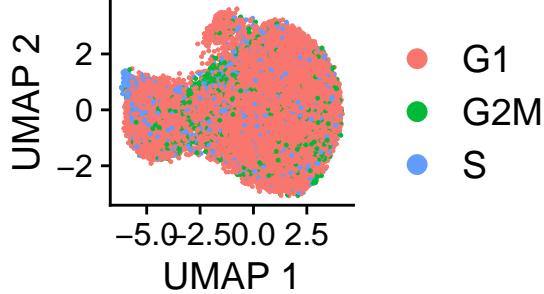
Biological Replicate



Perturbation Status



Cell Cycle Phase



Run mixscape

```
# Run mixscape.
eccite <- RunMixscape(
  object = eccite,
  assay = "PRTB",
  slot = "scale.data",
  labels = "gene",
  nt.class.name = "NT",
  min.de.genes = 5,
  iter.num = 10,
  de.assay = "RNA",
  verbose = F,
  prtb.type = "KO")

## Warning in FindMarkers.default(object = data.use, slot = data.slot, counts =
## counts, : No features pass logfc.threshold threshold; returning empty
## data.frame

## number of iterations= 95
## number of iterations= 187
## number of iterations= 172
```

```

## number of iterations= 18
## number of iterations= 6
## number of iterations= 18
## number of iterations= 11
## number of iterations= 11
## number of iterations= 59
## number of iterations= 43
## number of iterations= 42
## number of iterations= 19
## number of iterations= 12
## number of iterations= 12
## number of iterations= 23
## number of iterations= 19
## number of iterations= 19
## number of iterations= 51
## number of iterations= 51
## number of iterations= 51
## number of iterations= 36
## number of iterations= 26
## number of iterations= 25
## number of iterations= 20
## number of iterations= 12
## number of iterations= 12
## number of iterations= 17
## number of iterations= 15
## number of iterations= 14
## number of iterations= 13
## number of iterations= 73
## number of iterations= 46
## number of iterations= 41

# Calculate percentage of KO cells for all target gene classes.
df <- prop.table(table(eccite$mixscape_class.global, eccite$NT),2)

df2 <- reshape2::melt(df)
df2$Var2 <- as.character(df2$Var2)
test <- df2[which(df2$Var1 == "KO"),]
test <- test[order(test$value, decreasing = T),]
new.levels <- test$Var2
df2$Var2 <- factor(df2$Var2, levels = new.levels )
df2$Var1 <- factor(df2$Var1, levels = c("NT", "NP", "KO"))
df2$gene <- sapply(as.character(df2$Var2), function(x) strsplit(x, split = "g")[[1]][1])
df2$guide_number <- sapply(as.character(df2$Var2),
                           function(x) strsplit(x, split = "g")[[1]][2])
df3 <- df2[-c(which(df2$gene == "NT")),]

p1 <- ggplot(df3, aes(x = guide_number, y = value*100, fill= Var1)) +
  geom_bar(stat= "identity") +
  theme_classic()+
  scale_fill_manual(values = c("grey49", "grey79","coral1")) +
  ylab("% of cells") +
  xlab("sgRNA")

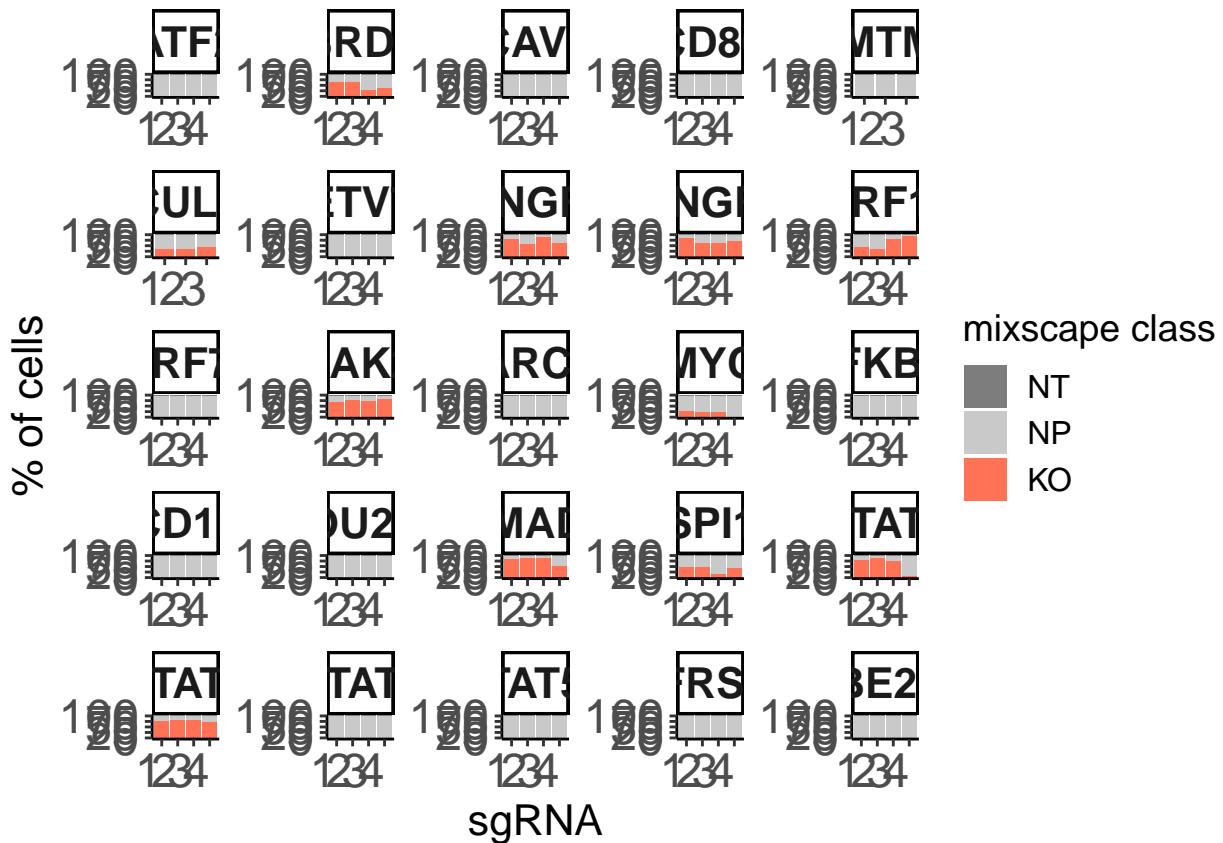
p1 + theme(axis.text.x = element_text(size = 18, hjust = 1),

```

```

axis.text.y = element_text(size = 18),
axis.title = element_text(size = 16),
strip.text = element_text(size=16, face = "bold")) +
facet_wrap(vars(gene), ncol = 5, scales = "free") +
labs(fill = "mixscape class") +theme(legend.title = element_text(size = 14),
legend.text = element_text(size = 12))

```



Perform Marker Gene Analysis/Pathway Enrichment Analysis Ordered by Log Fold Change

Papalex et reports 5 major pathways in CUL3 KO cells. Nuclear Receptors Meta-Pathway,NRF2 pathway,Phytochemical activity on NRF2 transcriptional activation,TYROBP causal network ,and complement system activation. We see pvalues that are approximately the same. However, we do see that the pathway “Phytochemical activity on NRF2 transcriptional activation” is ranked ahead of “TYROBP causal network” in this analysis whereas according to paplexi et al, the relative pvalue rank should be flipped between these two pathways.

```
CUL3_marker = FindMarkers(eccite,ident.1 = 'CUL3 KO',ident.2 = 'NT',
                           assay = 'RNA',only.pos = T)
```

```
## For a more efficient implementation of the Wilcoxon Rank Sum Test,
## (default method for FindMarkers) please install the limma package
```

```

## -----
## install.packages('BiocManager')
## BiocManager::install('limma')
## -----
## After installation of limma, Seurat will automatically use the more
## efficient implementation (no further action necessary).
## This message will be shown once per session

CUL3_marker$avg_log2FC = signif(CUL3_marker$avg_log2FC, digits=2)
CUL3_marker$p_val = signif(CUL3_marker$p_val, digits=2)

#get top 300 genes by log fold change
top = 300
CUL3_top = rownames(CUL3_marker)[order(CUL3_marker$avg_log2FC,decreasing = T
                                         )[1:top]]
#run pathway enrichment analysis
pathways = enrichr(CUL3_top,databases = 'WikiPathway_2021_Human')

## Uploading data to Enrichr... Done.
## Querying WikiPathway_2021_Human... Done.
## Parsing results... Done.

#get term and pvalue (using subset truncates pvalues for some reason)
pathways_seurat_log = cbind(pathways$WikiPathway_2021_Human$Term,
                             pathways$WikiPathway_2021_Human$Adjusted.P.value)[1:50,]
colnames(pathways_seurat_log) = c('Term','Adj.Pvalue')

pathways_seurat_log[,2] = signif(as.numeric(pathways_seurat_log[,2]), digits=2)
#make table
results = kable(pathways_seurat_log,booktabs = TRUE, linesep = "",
                caption = "Seurat Pathway Enrichment Analysis With Top 300 Genes by Log Fold
Change")
kable_styling(results,position = "center", latex_options = "scale_down")

```

Table 1: Seurat Pathway Enrichment Analysis With Top 300 Genes by Log Fold Change

Term	Adj.Pvalue
Nuclear Receptors Meta-Pathway WP2882	1.2e-19
IL-18 signaling pathway WP4754	2.7e-15
NRF2 pathway WP2884	5.8e-14
Lung fibrosis WP3624	7.7e-10
Spinal Cord Injury WP2431	2.1e-08
Phytochemical activity on NRF2 transcriptional activation WP3	5.9e-08
SARS-CoV-2 innate immunity evasion and cell-specific immune response WP5039	2e-07
Ferroptosis WP4313	3e-07
Selenium Micronutrient Network WP15	3.8e-07
TYROBP causal network in microglia WP3945	8.2e-07
Regulation of toll-like receptor signaling pathway WP1449	8.2e-07
Chemokine signaling pathway WP3929	8.8e-07
NRF2-ARE regulation WP4357	9.4e-07
IL1 and megakaryocytes in obesity WP2865	1.1e-06
Photodynamic therapy-induced NFE2L2 (NRF2) survival signaling WP3612	1.1e-06
Toll-like Receptor Signaling Pathway WP75	1.1e-06
Photodynamic therapy-induced AP-1 survival signaling. WP3611	1.1e-06
p53 transcriptional gene network WP4963	1.2e-06
Senescence and Autophagy in Cancer WP615	1.2e-06
Apoptosis-related network due to altered Notch3 in ovarian cancer WP2864	1.6e-06
Allograft Rejection WP2328	1.7e-06
VEGFA-VEGFR2 Signaling Pathway WP3888	3e-06
Oxidative Stress WP408	8.2e-06
Vitamin D Receptor Pathway WP2877	1.1e-05
Vitamin B12 metabolism WP1533	1.1e-05
Photodynamic therapy-induced NF-kB survival signaling WP3617	1.1e-05
Pentose Phosphate Metabolism WP134	2.4e-05
Microglia Pathogen Phagocytosis Pathway WP3937	2.5e-05
Oxidative Damage WP3941	2.5e-05
COVID-19 adverse outcome pathway WP4891	2.5e-05
Complement and Coagulation Cascades WP558	2.9e-05
Platelet-mediated interactions with vascular and circulating cells WP4462	4.7e-05
Aryl Hydrocarbon Receptor Netpath WP2586	6e-05
Glucocorticoid Receptor Pathway WP2880	0.00011
Complement system WP2806	0.00017
Kynurenone Pathway and links to Cellular Senescence WP5044	0.00021
Fibrin Complement Receptor 3 Signaling Pathway WP4136	0.00032
Vitamin D-sensitive calcium signaling in depression WP4698	0.00032
Cytokines and Inflammatory Response WP530	0.00037
Folate Metabolism WP176	0.00074
Tryptophan catabolism leading to NAD+ production WP4210	0.00074
RANKL/RANK signaling pathway WP2018	0.0016
miRNAs involvement in the immune response in sepsis WP4329	0.0019
Photodynamic therapy-induced HIF-1 survival signaling WP3614	0.0019
Complement Activation WP545	0.0025
Ebstein-Barr virus LMP1 signaling WP262	0.0029
Glutathione metabolism WP100	0.0029
Metabolic reprogramming in colon cancer WP4290	0.0032
NAD Metabolism in Oncogene-Induced Senescence and Mitochondrial Dysfunction-Associated Senescence WP5046	0.0032
Unfolded protein response WP4925	0.0032

Perform Marker Gene Analysis/Pathway Enrichment Analysis By Pvalue

Sorting by pvalue returns similar results but the pvalues are not as extreme. However, the top 5 processes reported are much close to the top 5 overall. Overall I would say the results are consistent with what was reported in the paper regardless of how you choose the top genes.

```
#get top 300 genes by pvalue
top = 300
CUL3_top = rownames(CUL3_marker)[order(CUL3_marker$p_val,decreasing = F)[1:top]]
#run pathway enrichment analysis
pathways = enrichr(CUL3_top,databases = 'WikiPathway_2021_Human')

## Uploading data to Enrichr... Done.
##   Querying WikiPathway_2021_Human... Done.
##   Parsing results... Done.

#get term and pvalue (using subset truncates pvalues for some reason)
pathways_seurat_pval = cbind(pathways$WikiPathway_2021_Human$Term,
                               pathways$WikiPathway_2021_Human$Adjusted.P.value)[1:50,]
pathways_seurat_pval[,2] = signif(as.numeric(pathways_seurat_pval[,2]),digits=2)
colnames(pathways_seurat_pval) = c('Term','Adj.Pvalue')
#make table
results = kable(pathways_seurat_pval,booktabs = TRUE, linesep = "",
                 caption = "Seurat Pathway Enrichment Analysis With Top 300 Genes by Pvalue")
kable_styling(results,position = "center", latex_options = "scale_down")
```

Table 2: Seurat Pathway Enrichment Analysis With Top 300 Genes by Pvalue

Term	Adj.Pvalue
Nuclear Receptors Meta-Pathway WP2882	5.4e-14
NRF2 pathway WP2884	1.2e-11
IL-18 signaling pathway WP4754	3e-09
TYROBP causal network in microglia WP3945	2.1e-06
Phytochemical activity on NRF2 transcriptional activation WP3	3e-06
SARS-CoV-2 innate immunity evasion and cell-specific immune response WP5039	3e-06
Selenium Micronutrient Network WP15	4.8e-06
Ferroptosis WP4313	4.8e-06
Oxidative Damage WP3941	4.8e-06
Complement system WP2806	8.3e-06
Lung fibrosis WP3624	1.3e-05
Senescence and Autophagy in Cancer WP615	1.6e-05
NRF2-ARE regulation WP4357	2.5e-05
Apoptosis-related network due to altered Notch3 in ovarian cancer WP2864	2.8e-05
Photodynamic therapy-induced NFE2L2 (NRF2) survival signaling WP3612	2.8e-05
Regulation of toll-like receptor signaling pathway WP1449	2.9e-05
COVID-19 adverse outcome pathway WP4891	4.2e-05
Toll-like Receptor Signaling Pathway WP75	6.9e-05
Aryl Hydrocarbon Receptor Netpath WP2586	9.9e-05
p53 transcriptional gene network WP4963	0.00013
Oxidative Stress WP408	0.00015
Spinal Cord Injury WP2431	0.00019
Photodynamic therapy-induced NF-kB survival signaling WP3617	0.00019
Complement Activation WP545	0.00024
Type II interferon signaling (IFNG) WP619	0.00025
Kynurenine Pathway and links to Cellular Senescence WP5044	0.00028
Complement and Coagulation Cascades WP558	0.00033
Microglia Pathogen Phagocytosis Pathway WP3937	0.00035
Vitamin D-sensitive calcium signaling in depression WP4698	0.00038
Fibrin Complement Receptor 3 Signaling Pathway WP4136	0.00038
Chemokine signaling pathway WP3929	0.00045
Allograft Rejection WP2328	0.00065
Aryl Hydrocarbon Receptor Pathway WP2873	0.00067
Folate Metabolism WP176	0.00083
Vitamin B12 metabolism WP1533	0.001
Platelet-mediated interactions with vascular and circulating cells WP4462	0.001
Pentose Phosphate Metabolism WP134	0.0011
Adipogenesis WP236	0.0015
Overview of leukocyte-intrinsic Hippo pathway functions WP4542	0.0015
RANKL/RANK signaling pathway WP2018	0.0015
miRNAs involvement in the immune response in sepsis WP4329	0.0019
FGF23 signaling in hypophosphatemic rickets and related disorders WP4790	0.0025
Glutathione metabolism WP100	0.003
Sphingolipid Metabolism (general overview) WP4725	0.0033
IL1 and megakaryocytes in obesity WP2865	0.0033
NAD Metabolism in Oncogene-Induced Senescence and Mitochondrial Dysfunction-Associated Senescence WP5046	0.0033
TNF-alpha signaling pathway WP231	0.0036
Sphingolipid Metabolism (integrated pathway) WP4726	0.0037
Cytokines and Inflammatory Response WP530	0.0042
Non-genomic actions of 1,25 dihydroxyvitamin D3 WP4341	0.005

SCEPTRE Pathway Analysis

```
#using absolute paths to download results since files exist on github
code_dir = .get_config_path("LOCAL_CODE_DIR")
data.dir = paste0(code_dir,"/sceptre2-manuscript/writeups/papalex_i_analysis/")
gene_path = paste0(data.dir,
                    'sceptre_CUL3_and_PDL1_mrna_results_with_effect_size.rds')

#Note that sceptre results have columns pvalue, grna, target
#get sceptre perturbation on PDL1 mrna results
gene_result = readRDS(gene_path)
#gene_result$log_fold_change = signif(gene_result$log_fold_change, digits=2)

top = 300
#Get CUL3 results
CUL3_SCEPTRE = subset(gene_result,grna_group == "CUL3")

#get top 300 genes by pvalue
pval_pos = CUL3_SCEPTRE$p_value*sign(CUL3_SCEPTRE$log_fold_change)
pval_pos[pval_pos < 0] = 1000000
order_pval = order(pval_pos)[1:top]
order_fold = order(CUL3_SCEPTRE$log_fold_change,decreasing = T)[1:top]
SCEPTRE_fold = unfactor(CUL3_SCEPTRE$response_id[order_fold])
SCEPTRE_pval = unfactor(CUL3_SCEPTRE$response_id[order_pval])
```

By Log Fold Change

We see essentially no overlap between SCEPTRE and Seurat.

```
pathways = enrichr(SCEPTRE_fold,databases = 'WikiPathway_2021_Human')

## Uploading data to Enrichr... Done.
##   Querying WikiPathway_2021_Human... Done.
## Parsing results... Done.

#get term and pvalue (using subset truncates pvalues for some reason)
pathways_sceptre_log = cbind(pathways$WikiPathway_2021_Human$Term,
                             pathways$WikiPathway_2021_Human$Adjusted.P.value)[1:50,]
pathways_sceptre_log[,2] = signif(as.numeric(pathways_sceptre_log[,2]),digits=2)
colnames(pathways_sceptre_log) = c('Term','Adj.Pvalue')
#make table
results = kable(pathways_sceptre_log,booktabs = TRUE, linesep = "",
                caption = "SCEPTRE Pathway Enrichment Analysis With Top 300 Genes by
Log Fold Change")
kable_styling(results,position = "center", latex_options = "scale_down")
```

Table 3: SCEPTRE Pathway Enrichment Analysis With Top 300 Genes by Log Fold Change

Term	Adj.Pvalue
IL-18 signaling pathway WP4754	0.0038
Lung fibrosis WP3624	0.005
Oligodendrocyte specification and differentiation, leading to myelin components for CNS WP4304	0.0059
Photodynamic therapy-induced NF- κ B survival signaling WP3617	0.0095
Serotonin Transporter Activity WP1455	0.023
Senescence and Autophagy in Cancer WP615	0.04
Signal transduction through IL1R WP4496	0.048
Small Ligand GPCRs WP247	0.065
Focal Adhesion WP306	0.065
SARS-CoV-2 innate immunity evasion and cell-specific immune response WP5039	0.065
Small cell lung cancer WP4658	0.065
Chemokine signaling pathway WP3929	0.065
Prostaglandin Synthesis and Regulation WP98	0.078
Kynurenine Pathway and links to Cellular Senescence WP5044	0.078
IL1 and megakaryocytes in obesity WP2865	0.082
Focal Adhesion-PI3K-Akt-mTOR-signaling pathway WP3932	0.091
Cytokines and Inflammatory Response WP530	0.091
Spinal Cord Injury WP2431	0.11
Selective expression of chemokine receptors during T-cell polarization WP4494	0.11
Allograft Rejection WP2328	0.12
Hair Follicle Development: Organogenesis - Part 2 of 3 WP2839	0.12
Monoamine Transport WP727	0.12
Ovarian infertility WP34	0.12
PI3K-Akt signaling pathway WP4172	0.14
miRNAs involvement in the immune response in sepsis WP4329	0.16
Neovascularisation processes WP4331	0.16
COVID-19 adverse outcome pathway WP4891	0.18
Tryptophan catabolism leading to NAD ⁺ production WP4210	0.19
Cytosolic DNA-sensing pathway WP4655	0.2
Platelet-mediated interactions with vascular and circulating cells WP4462	0.2
Vitamin A and carotenoid metabolism WP716	0.2
Simplified Interaction Map Between LOXL4 and Oxidative Stress Pathway WP3670	0.21
LTf danger signal response pathway WP4478	0.23
Differentiation Pathway WP2848	0.24
Hypertrophy Model WP516	0.24
Netrin-UNC5B signaling pathway WP4747	0.26
Selenium Micronutrient Network WP15	0.26
Apoptosis-related network due to altered Notch3 in ovarian cancer WP2864	0.26
Cardiac Progenitor Differentiation WP2406	0.26
Ebstein-Barr virus LMP1 signaling WP262	0.26
TGF-beta Receptor Signaling WP560	0.26
Hematopoietic Stem Cell Differentiation WP2849	0.26
IL-1 signaling pathway WP195	0.26
NAD Metabolism in Oncogene-Induced Senescence and Mitochondrial Dysfunction-Associated Senescence WP5046	0.26
Relationship between inflammation, COX-2 and EGFR WP4483	0.28
TGF-beta receptor signaling in skeletal dysplasias WP4816	0.28
RAS and bradykinin pathways in COVID-19 WP4969	0.34
Dopaminergic Neurogenesis WP2855	0.34
Eicosanoid metabolism via cyclooxygenases (COX) WP4719	0.34
Extracellular vesicle-mediated signaling in recipient cells WP2870	0.34

SCEPTRE Pathway Enrichment By Pvalue

Sorting by Pvalue gives more similar results to Seurat (NRF2,Nuclear Receptors,Phytochemical Activity), but it is missing the complement system and the TYROBP causal network. However, since NRF2 is the main topic of their paper, the differences do not have very large consequences.

```
pathways = enrichr(SCEPTRE_pval,databases = 'WikiPathway_2021_Human')

## Uploading data to Enrichr... Done.
##   Querying WikiPathway_2021_Human... Done.
##   Parsing results... Done.

#get term and pvalue (using subset truncates pvalues for some reason)
pathways_sceptre_pval = cbind(pathways$WikiPathway_2021_Human$Term,
                               pathways$WikiPathway_2021_Human$Adjusted.P.value)[1:50,]
pathways_sceptre_pval[,2] = signif(as.numeric(pathways_sceptre_pval[,2]),
                                   digits=2)
colnames(pathways_sceptre_pval) = c('Term','Adj.Pvalue')
#make table
results = kable(pathways_sceptre_pval,booktabs = TRUE, linesep = "",
                caption = "SCEPTRE Pathway Enrichment Analysis With Top 300 Genes by Pvalue")
kable_styling(results,position = "center", latex_options = "scale_down")
```

Table 4: SCEPTRE Pathway Enrichment Analysis With Top 300 Genes by Pvalue

Term	Adj.Pvalue
Nuclear Receptors Meta-Pathway WP2882	1e-18
NRF2 pathway WP2884	4.3e-16
IL-18 signaling pathway WP4754	6e-11
Selenium Micronutrient Network WP15	6.8e-08
Phytochemical activity on NRF2 transcriptional activation WP3	6.8e-08
Apoptosis-related network due to altered Notch3 in ovarian cancer WP2864	3.3e-07
NRF2-ARE regulation WP4357	1.7e-06
Photodynamic therapy-induced NFE2L2 (NRF2) survival signaling WP3612	2e-06
Aryl Hydrocarbon Receptor Netpath WP2586	1.5e-05
Oxidative Stress WP408	1.8e-05
SARS-CoV-2 innate immunity evasion and cell-specific immune response WP5039	2e-05
Photodynamic therapy-induced NF- κ B survival signaling WP3617	2.3e-05
Pentose Phosphate Metabolism WP134	4.8e-05
Ferroptosis WP4313	4.8e-05
COVID-19 adverse outcome pathway WP4891	4.8e-05
Fibrin Complement Receptor 3 Signaling Pathway WP4136	5.3e-05
Chemokine signaling pathway WP3929	0.00015
Regulation of toll-like receptor signaling pathway WP1449	0.00017
Folate Metabolism WP176	0.00017
Vitamin B12 metabolism WP1533	0.00017
Type II interferon signaling (IFNG) WP619	0.00029
RANKL/RANK signaling pathway WP2018	0.00029
Glutathione metabolism WP100	0.00031
Kynurene Pathway and links to Cellular Senescence WP5044	0.00031
Toll-like Receptor Signaling Pathway WP75	0.00035
IL1 and megakaryocytes in obesity WP2865	0.00035
Oxidative Damage WP3941	0.00037
Senescence and Autophagy in Cancer WP615	0.00038
VEGFA-VEGFR2 Signaling Pathway WP3888	0.00039
Vitamin D-sensitive calcium signaling in depression WP4698	0.00039
Lung fibrosis WP3624	0.00051
Aryl Hydrocarbon Receptor Pathway WP2873	0.00071
p53 transcriptional gene network WP4963	0.00072
Spinal Cord Injury WP2431	0.00077
TNF-alpha signaling pathway WP231	0.00077
Oligodendrocyte specification and differentiation, leading to myelin components for CNS WP4304	8e-04
Glucocorticoid Receptor Pathway WP2880	0.00085
Photodynamic therapy-induced AP-1 survival signaling. WP3611	0.00097
Platelet-mediated interactions with vascular and circulating cells WP4462	0.00097
Nanomaterial-induced inflammasome activation WP3890	0.001
Adipogenesis WP236	0.0014
LTf danger signal response pathway WP4478	0.0014
miRNAs involvement in the immune response in sepsis WP4329	0.0019
Role of Altered Glycolysation of MUC1 in Tumour Microenvironment WP4480	0.0022
Apoptosis WP254	0.0022
Ebstein-Barr virus LMP1 signaling WP262	0.0028
Allograft Rejection WP2328	0.003
TNF related weak inducer of apoptosis (TWEAK) Signaling Pathway WP2036	0.003
Metabolic reprogramming in colon cancer WP4290	0.003
NAD metabolism, sirtuins and aging WP3630	0.0038

Table 5: Seurat and SCEPTRÉ Top 300 Genes Overlap (Out of 300)

	Sceptre By Pvalue	Sceptre By log fold change
Seurat By Pvalue	202	40
Seurat By log fold change	196	35

Looking Into the Overlapped Genes

Clearly there is something going on when using log fold change with SCEPTRÉ's estimates. Looking at the results, we see that Seurat is consistent between ordering by pvalue and log fold change whereas SCEPTRÉ is not very consistent.

```
#get top 300 genes from seurat
seurat_pval = rownames(CUL3_marker)[order(CUL3_marker$p_val,decreasing = F)
[1:top]]
seurat_fold = rownames(CUL3_marker)[order(CUL3_marker$avg_log2FC,decreasing = T
)[1:top]]

overlap_mat = matrix(NA,2,2)
colnames(overlap_mat) = c('Sceptre By Pvalue','Sceptre By log fold change')
rownames(overlap_mat) = c('Seurat By Pvalue','Seurat By log fold change')

#see seurat by pvalue overlap with sceptre
overlap_mat[1,1] = sum(seurat_pval%in%SCEPTRÉ_pval)
overlap_mat[1,2] = sum(seurat_pval%in%SCEPTRÉ_fold)
#get seurat by LFC overlap woth sceptre
overlap_mat[2,1] = sum(seurat_fold%in%SCEPTRÉ_pval)
overlap_mat[2,2]= sum(seurat_fold%in%SCEPTRÉ_fold)

#get within method overlap
overlap_within = matrix(NA,1,2)
colnames(overlap_within) = c('seurat','SCEPTRÉ')
rownames(overlap_within) = c('Within Method Overlap')
overlap_within[,1] = sum(seurat_fold%in%seurat_pval)
overlap_within[,2] = sum(SCEPTRÉ_fold%in%SCEPTRÉ_pval)

#make table
results = kable(overlap_mat,booktabs = TRUE, linesep = "",
caption = "Seurat and SCEPTRÉ Top 300 Genes Overlap (Out of 300)")
kable_styling(results,position = "center", latex_options = "scale_down")

#make table
results = kable(overlap_within,booktabs = TRUE, linesep = "",
caption = "Seurat and SCEPTRÉ Top 300 Genes Overlap Within Each Method
(Out of 300)")
kable_styling(results,position = "center", latex_options = "scale_down")
```

Table 6: Seurat and SCEPTRE Top 300 Genes Overlap Within Each Method (Out of 300)

	seurat	SCEPTRE
Within Method Overlap	237	74

Checking if Results Are More Similar When Not Using All Significant Genes

Using all significant genes with a BH correction results in 321 genes when setting the cutoff to 0.05. The results are quite similar to those when ordering by pvalue. Something may be going on with SCEPTRE's large effect size estimates.

```
#get top 300 genes by pvalue
pval_pos = CUL3_SCEPTRE$p_value
#apply BH
pval_pos = p.adjust(pval_pos,method = 'BH')
#get which are significant
sig_genes = which(CUL3_SCEPTRE$log_fold_change > 0 & pval_pos < 0.05)
sig_genes = unfactor(CUL3_SCEPTRE$response_id[sig_genes])

pathways = enrichr(sig_genes,databases = 'WikiPathway_2021_Human')

## Uploading data to Enrichr... Done.
## Querying WikiPathway_2021_Human... Done.
## Parsing results... Done.

#get term and pvalue (using subset truncates pvalues for some reason)
pathways_sceptre_pval = cbind(pathways$WikiPathway_2021_Human$Term,
                               pathways$WikiPathway_2021_Human$Adjusted.P.value)[1:50,]
pathways_sceptre_pval[,2] = signif(as.numeric(pathways_sceptre_pval[,2]),
                                   digits=2)
colnames(pathways_sceptre_pval) = c('Term','Adj.Pvalue')
#make table
results = kable(pathways_sceptre_pval,booktabs = TRUE, linesep = "",
                caption = "SCEPTRE Pathway Enrichment Analysis With All Significant Genes")
kable_styling(results,position = "center", latex_options = "scale_down")
```

Table 7: SCEPTRE Pathway Enrichment Analysis With All Significant Genes

Term	Adj.Pvalue
Nuclear Receptors Meta-Pathway WP2882	1e-17
NRF2 pathway WP2884	2.1e-15
IL-18 signaling pathway WP4754	2.7e-10
Phytochemical activity on NRF2 transcriptional activation WP3	1.3e-07
Selenium Micronutrient Network WP15	1.3e-07
Apoptosis-related network due to altered Notch3 in ovarian cancer WP2864	6.5e-07
NRF2-ARE regulation WP4357	2.7e-06
Photodynamic therapy-induced NFE2L2 (NRF2) survival signaling WP3612	3.3e-06
SARS-CoV-2 innate immunity evasion and cell-specific immune response WP5039	3.9e-06
Aryl Hydrocarbon Receptor Netpath WP2586	2.3e-05
Oxidative Stress WP408	2.6e-05
Photodynamic therapy-induced NF-kB survival signaling WP3617	3.7e-05
Pentose Phosphate Metabolism WP134	6.4e-05
Chemokine signaling pathway WP3929	6.8e-05
COVID-19 adverse outcome pathway WP4891	6.8e-05
Ferroptosis WP4313	7.1e-05
Fibrin Complement Receptor 3 Signaling Pathway WP4136	8e-05
Vitamin B12 metabolism WP1533	0.00028
Folate Metabolism WP176	0.00028
Regulation of toll-like receptor signaling pathway WP1449	0.00029
Type II interferon signaling (IFNG) WP619	0.00043
RANKL/RANK signaling pathway WP2018	0.00043
Glutathione metabolism WP100	0.00043
Kynurene Pathway and links to Cellular Senescence WP5044	0.00043
IL1 and megakaryocytes in obesity WP2865	0.00052
Oxidative Damage WP3941	0.00057
Toll-like Receptor Signaling Pathway WP75	0.00059
Vitamin D-sensitive calcium signaling in depression WP4698	0.00062
Senescence and Autophagy in Cancer WP615	0.00064
Lung fibrosis WP3624	0.00083
VEGFA-VEGFR2 Signaling Pathway WP3888	0.00093
Aryl Hydrocarbon Receptor Pathway WP2873	0.001
Allograft Rejection WP2328	0.0011
p53 transcriptional gene network WP4963	0.0011
Oligodendrocyte specification and differentiation, leading to myelin components for CNS WP4304	0.0011
TNF-alpha signaling pathway WP231	0.0012
Spinal Cord Injury WP2431	0.0012
Glucocorticoid Receptor Pathway WP2880	0.0013
Platelet-mediated interactions with vascular and circulating cells WP4462	0.0013
Nanomaterial-induced inflammasome activation WP3890	0.0013
Photodynamic therapy-induced AP-1 survival signaling. WP3611	0.0013
LTF danger signal response pathway WP4478	0.0019
Adipogenesis WP236	0.0022
miRNAs involvement in the immune response in sepsis WP4329	0.0025
Role of Altered Glycolysation of MUC1 in Tumour Microenvironment WP4480	0.0027
Novel intracellular components of RIG-I-like receptor (RLR) pathway WP3865	0.0032
Apoptosis WP254	0.0032
Ebstein-Barr virus LMP1 signaling WP262	0.0036
TNF related weak inducer of apoptosis (TWEAK) Signaling Pathway WP2036	0.0041
Metabolic reprogramming in colon cancer WP4290	0.0041

Analyzing Number of Genes That Contribute to Processes

We know that SCEPTRE and Seurat Differ on the processes TYROBP and Photochemical activity. Looking at the table below we see that 10/61 and 7/15 of the genes contribute the significant pvalue seen when using Seurat. This is quite noisy. On the other hand, for NRF2, the overlap is 22/146 which is less noisy.

```
#get top 300 genes by log fold change
top = 300
CUL3_top = rownames(CUL3_marker)[order(CUL3_marker$avg_log2FC,decreasing = T
                                         )[1:top]]
#run pathway enrichment analysis
pathways = enrichr(CUL3_top,databases = 'WikiPathway_2021_Human')

## Uploading data to Enrichr... Done.
##   Querying WikiPathway_2021_Human... Done.
##   Parsing results... Done.

#get term and pvalue (using subset truncates pvalues for some reason)
pathways = cbind(pathways$WikiPathway_2021_Human$Term,
                  pathways$WikiPathway_2021_Human$Adjusted.P.value,
                  pathways$WikiPathway_2021_Human$Overlap)[1:50,]

pathways[,2] = signif(as.numeric(pathways_seurat_log[,2]), digits=2)
colnames(pathways) = c('Term','Adj.Pvalue','Overlap')

#make table
results = kable(pathways,booktabs = TRUE, linesep = "",
                 caption = "Seurat Pathway Enrichment Analysis With Top 300 Genes by Log Fold
Change")
kable_styling(results,position = "center", latex_options = "scale_down")
```

Table 8: Seurat Pathway Enrichment Analysis With Top 300 Genes by Log Fold Change

Term	Adj.Pvalue	Overlap
Nuclear Receptors Meta-Pathway WP2882	1.2e-19	37/319
IL-18 signaling pathway WP4754	2.7e-15	30/272
NRF2 pathway WP2884	5.8e-14	22/146
Lung fibrosis WP3624	7.7e-10	13/63
Spinal Cord Injury WP2431	2.1e-08	15/118
Phytochemical activity on NRF2 transcriptional activation WP3	5.9e-08	7/15
SARS-CoV-2 innate immunity evasion and cell-specific immune response WP5039	2e-07	11/66
Ferroptosis WP4313	3e-07	9/40
Selenium Micronutrient Network WP15	3.8e-07	12/89
TYROBP causal network in microglia WP3945	8.2e-07	10/61
Regulation of toll-like receptor signaling pathway WP1449	8.2e-07	14/139
Chemokine signaling pathway WP3929	8.8e-07	15/164
NRF2-ARE regulation WP4357	9.4e-07	7/23
IL1 and megakaryocytes in obesity WP2865	1.1e-06	7/24
Photodynamic therapy-induced NFE2L2 (NRF2) survival signaling WP3612	1.1e-06	7/24
Toll-like Receptor Signaling Pathway WP75	1.1e-06	12/103
Photodynamic therapy-induced AP-1 survival signaling. WP3611	1.1e-06	9/50
p53 transcriptional gene network WP4963	1.2e-06	10/67
Senescence and Autophagy in Cancer WP615	1.2e-06	12/105
Apoptosis-related network due to altered Notch3 in ovarian cancer WP2864	1.6e-06	9/53
Allograft Rejection WP2328	1.7e-06	11/89
VEGFA-VEGFR2 Signaling Pathway WP3888	3e-06	23/432
Oxidative Stress WP408	8.2e-06	7/33
Vitamin D Receptor Pathway WP2877	1.1e-05	14/182
Vitamin B12 metabolism WP1533	1.1e-05	8/50
Photodynamic therapy-induced NF- κ B survival signaling WP3617	1.1e-05	7/35
Pentose Phosphate Metabolism WP134	2.4e-05	4/7
Microglia Pathogen Phagocytosis Pathway WP3937	2.5e-05	7/40
Oxidative Damage WP3941	2.5e-05	7/40
COVID-19 adverse outcome pathway WP4891	2.5e-05	5/15
Complement and Coagulation Cascades WP558	2.9e-05	8/58
Platelet-mediated interactions with vascular and circulating cells WP4462	4.7e-05	5/17
Aryl Hydrocarbon Receptor Netpath WP2586	6e-05	7/46
Glucocorticoid Receptor Pathway WP2880	0.00011	8/70
Complement system WP2806	0.00017	9/97
Kynurenone Pathway and links to Cellular Senescence WP5044	0.00021	5/23
Fibrin Complement Receptor 3 Signaling Pathway WP4136	0.00032	6/41
Vitamin D-sensitive calcium signaling in depression WP4698	0.00032	6/41
Cytokines and Inflammatory Response WP530	0.00037	5/26
Folate Metabolism WP176	0.00074	7/69
Tryptophan catabolism leading to NAD+ production WP4210	0.00074	4/16
RANKL/RANK signaling pathway WP2018	0.0016	6/55
miRNAs involvement in the immune response in sepsis WP4329	0.0019	5/37
Photodynamic therapy-induced HIF-1 survival signaling WP3614	0.0019	5/37
Complement Activation WP545	0.0025	4/22
Ebstein-Barr virus LMP1 signaling WP262	0.0029	4/23
Glutathione metabolism WP100	0.0029	4/23
Metabolic reprogramming in colon cancer WP4290	0.0032	5/42
NAD Metabolism in Oncogene-Induced Senescence and Mitochondrial Dysfunction-Associated Senescence WP5046	0.0032	4/24
Unfolded protein response WP4925	0.0032	4/24

Checking If Seurat Filters Drive Differences

Seurat has two filters that change what genes are tested. Logfc.threshold makes it so that only genes that show a log fold change greater than some level are ouputted. The default value is 0.25. The min.pct argument makes it so that only genes that are present in over a certain percentage of cells in each class are tested. We see that when including all genes, both Seurat and SCEPTR have little correspondence with both each other and with themselves.

```
CUL3_marker_all = FindMarkers(eccite,ident.1 = 'CUL3 KO',ident.2 = 'NT',
                               assay = 'RNA',only.pos = F,logfc.threshold = 0,min.pct = 0)
CUL3_marker_all$avg_log2FC = signif(CUL3_marker_all$avg_log2FC, digits=2)
CUL3_marker_all$p_val = signif(CUL3_marker_all$p_val, digits=2)
```

```
#get top seurat genes by pvalue and log fold change
seurat_fold_all = rownames(CUL3_marker_all)[
  order(CUL3_marker_all$avg_log2FC,decreasing = T)[1:top]]
P_seurat = CUL3_marker_all$p_val*sign(CUL3_marker_all$avg_log2FC)
P_seurat [P_seurat < 0] = 1000000
seurat_pval_all = rownames(CUL3_marker_all)[order(pval_pos)[1:top]]

#
```

```
overlap_mat = matrix(NA,2,2)
colnames(overlap_mat) = c('Sceptre By Pvalue','Sceptre By log fold change')
rownames(overlap_mat) = c('Seurat By Pvalue','Seurat By log fold change')

#see seurat by pvalue overlap with sceptre
overlap_mat[1,1] = sum(seurat_pval_all%in%SCEPTRE_pval)
overlap_mat[1,2] = sum(seurat_pval_all%in%SCEPTRE_fold)
#get seurat by LFC overlap woth sceptre
overlap_mat[2,1] = sum(seurat_fold_all%in%SCEPTRE_pval)
overlap_mat[2,2]= sum(seurat_fold_all%in%SCEPTRE_fold)

#get within method overlap
overlap_within = matrix(NA,1,2)
colnames(overlap_within) = c('seurat','SCEPTRE')
rownames(overlap_within) = c('Within Method Overlap')
overlap_within[,1] = sum(seurat_fold_all%in%seurat_pval_all)
overlap_within[,2] = sum(SCEPTRE_fold%in%SCEPTRE_pval)

#make table
results = kable(overlap_mat,booktabs = TRUE, linesep = "",
                caption = "Unfiltered Seurat and SCEPTR Top 300 Genes Overlap (Out of 300)")
kable_styling(results,position = "center", latex_options = "scale_down")
```

```
#make table
results = kable(overlap_within,booktabs = TRUE, linesep = "",
                caption = "Unfiltered Seurat and SCEPTR Top 300 Genes Overlap Within Each
Method (Out of 300)")
kable_styling(results,position = "center", latex_options = "scale_down")
```

Table 9: Unfiltered Seurat and SCEPTRE Top 300 Genes Overlap (Out of 300)

	Sceptre By Pvalue	Sceptre By log fold change
Seurat By Pvalue	6	8
Seurat By log fold change	197	38

Table 10: Unfiltered Seurat and SCEPTRE Top 300 Genes Overlap Within Each Method (Out of 300)

	seurat	SCEPTRE
Within Method Overlap	6	74

Checking Correspondence When Filtering SCEPTRE

```
top = 300
#Get CUL3 results
CUL3_SCEPTRE = subset(gene_result, grna_group == "CUL3")

#get top 300 genes by pvalue
pval_pos = CUL3_SCEPTRE$p_value*sign(CUL3_SCEPTRE$log_fold_change)
pval_pos[pval_pos < 0 | CUL3_SCEPTRE$log_fold_change > 0.25] = 1000000
order_pval = order(pval_pos)[1:top]
order_pval = order_pval[pval_pos[order_pval] < 1]
order_fold = order(CUL3_SCEPTRE$log_fold_change, decreasing = T)[1:top]
SCEPTRE_fold = unfactor(CUL3_SCEPTRE$response_id[order_fold])
SCEPTRE_pval = unfactor(CUL3_SCEPTRE$response_id[order_pval])
```

SCEPTRE Pathway Enrichment By Pvalue

Using SCEPTRE ordered by pvalue removes a lot of the shared signal detected by SCEPTRE and Seurat. We see the NRF2 is barely significant after pvalue adjustment and that the Nuclear Receptors Meta-Pathway has an adjusted pvalue of 0.1. The TYROBP pathway is now the most significant process.

```
pathways = enrichr(SCEPTRE_pval, databases = 'WikiPathway_2021_Human')
```

```
## Uploading data to Enrichr... Done.
##   Querying WikiPathway_2021_Human... Done.
##   Parsing results... Done.

#get term and pvalue (using subset truncates pvalues for some reason)
pathways_sceptre_pval = cbind(pathways$WikiPathway_2021_Human$Term,
                               pathways$WikiPathway_2021_Human$Adjusted.P.value)[1:50,]
pathways_sceptre_pval[,2] = signif(as.numeric(pathways_sceptre_pval[,2]),
```

```
          digits=2)
colnames(pathways_sceptre_pval) = c('Term','Adj.Pvalue')
#make table
results = kable(pathways_sceptre_pval,booktabs = TRUE, linesep = "",
caption = "Filtered SCEPTRE Pathway Analysis With Top 300 Genes by Pvalue")
kable_styling(results,position = "center", latex_options = "scale_down")
```

Table 11: Filtered SCEPTRE Pathway Analysis With Top 300 Genes by Pvalue

Term	Adj.Pvalue
VEGFA-VEGFR2 Signaling Pathway WP3888	0.00089
Photodynamic therapy-induced AP-1 survival signaling. WP3611	0.0016
Hepatitis B infection WP4666	0.011
Overview of leukocyte-intrinsic Hippo pathway functions WP4542	0.011
Photodynamic therapy-induced NF-kB survival signaling WP3617	0.011
Senescence and Autophagy in Cancer WP615	0.011
IL-18 signaling pathway WP4754	0.012
Apoptosis WP254	0.012
SARS-CoV-2 innate immunity evasion and cell-specific immune response WP5039	0.016
TNF-alpha signaling pathway WP231	0.016
Ebola Virus Pathway on Host WP4217	0.023
Copper homeostasis WP3286	0.03
Toll-like Receptor Signaling related to MyD88 WP3858	0.03
Interferon type I signaling pathways WP585	0.031
RANKL/RANK signaling pathway WP2018	0.032
NRF2 pathway WP2884	0.036
TYROBP causal network in microglia WP3945	0.043
Photodynamic therapy-induced HIF-1 survival signaling WP3614	0.043
Apoptosis Modulation and Signaling WP1772	0.043
TP53 network WP1742	0.047
Vitamin D in inflammatory diseases WP4482	0.069
Toll-like Receptor Signaling Pathway WP75	0.072
Head and Neck Squamous Cell Carcinoma WP4674	0.072
Unfolded protein response WP4925	0.074
Photodynamic therapy-induced NFE2L2 (NRF2) survival signaling WP3612	0.074
Exercise-induced Circadian Regulation WP410	0.076
ApoE and miR-146 in inflammation and atherosclerosis WP3926	0.076
Structural Pathway of Interleukin 1 (IL-1) WP2637	0.076
Thiamine metabolic pathways WP4297	0.09
Nuclear Receptors Meta-Pathway WP2882	0.1
Pathogenic Escherichia coli infection WP2272	0.1
Acute viral myocarditis WP4298	0.1
Mitochondrial complex I assembly model OXPHOS system WP4324	0.1
Insulin Signaling WP481	0.1
Selenium Micronutrient Network WP15	0.1
SARS coronavirus and innate immunity WP4912	0.1
SARS-CoV-2 mitochondrial interactions WP5038	0.1
Chemokine signaling pathway WP3929	0.1
MAPK Signaling Pathway WP382	0.1
Novel intracellular components of RIG-I-like receptor (RLR) pathway WP3865	0.1
Oxidative phosphorylation WP623	0.1
Corticotropin-releasing hormone signaling pathway WP2355	0.1
Host-pathogen interaction of human coronaviruses - interferon induction WP4880	0.1
Notch Signaling Pathway Netpath WP61	0.1
Computational Model of Aerobic Glycolysis WP4629	0.1
Proteasome Degradation WP183	0.1
CAMKK2 Pathway WP4874	0.1
TGF-beta Signaling Pathway WP366	0.11
Host-pathogen interaction of human coronaviruses ²⁸ - MAPK signaling WP4877	0.12
AGE/RAGE pathway WP2324	0.12

Table 12: Seurat and Filtered SCEPTRE Top 300 Genes Overlap (Out of 300)

	Sceptre By Pvalue	Sceptre By log fold change
Seurat By Pvalue	74	40
Seurat By log fold change	81	35

Additionally, there is even less of a correspondence between Seurat and SCEPTRE.

```

overlap_mat = matrix(NA,2,2)
colnames(overlap_mat) = c('Sceptre By Pvalue','Sceptre By log fold change')
rownames(overlap_mat) = c('Seurat By Pvalue','Seurat By log fold change')

#see seurat by pvalue overlap with sceptre
overlap_mat[1,1] = sum(seurat_pval%in%SCEPTRE_pval)
overlap_mat[1,2] = sum(seurat_pval%in%SCEPTRE_fold)
#get seurat by LFC overlap woth sceptre
overlap_mat[2,1] = sum(seurat_fold%in%SCEPTRE_pval)
overlap_mat[2,2]= sum(seurat_fold%in%SCEPTRE_fold)

#get within method overlap
overlap_within = matrix(NA,1,2)
colnames(overlap_within) = c('seurat','SCEPTRE')
rownames(overlap_within) = c('Within Method Overlap')
overlap_within[,1] = sum(seurat_fold%in%seurat_pval)
overlap_within[,2] = sum(SCEPTRE_fold%in%SCEPTRE_pval)

#make table
results = kable(overlap_mat,booktabs = TRUE, linesep = "",
caption = "Seurat and Filtered SCEPTRE Top 300 Genes Overlap (Out of 300)")
kable_styling(results,position = "center", latex_options = "scale_down")

```