MEDICAL MICROBIOLOGY AND INFECTIOUS DISEASES CODING WORKSHOP

Presents

Data visualization using antiviral drug-repurposing results from CLUE

INSTRUCTED BY

MacKenzie Wilke (MSc Student)

Email: sarvism@myumanitoba.ca

INFORMATION FOR PARTICIPANTS

All workshops are being recorded and posted to the MMID Coding Workshop - YouTube

Question and Answer period will not be recorded.

Last 3 Workshops...

- Introduction to R (Rstudio, R scripts, R markdown)
  - Grace E. Seo
- Tidy data in R (Transforming and combining data)
  - Molly Pratt
- Visualizing data using R (ggplot2)
  - Samantha Lee

LEARNING OBJECTIVES

1. *Understand what drug repurposing analyses are and how to use it as a tool for host-pathogen data*
2. *Create a basic plot in R to visualize CLUE results*
3. *Refine R plot components for specific colours, shapes and sizes*
4. *Combine multiple plots in R and output as a file to the desktop*
5. *Learn how to produce a similar graph in Tableau*

LEARNING OBJECTIVES

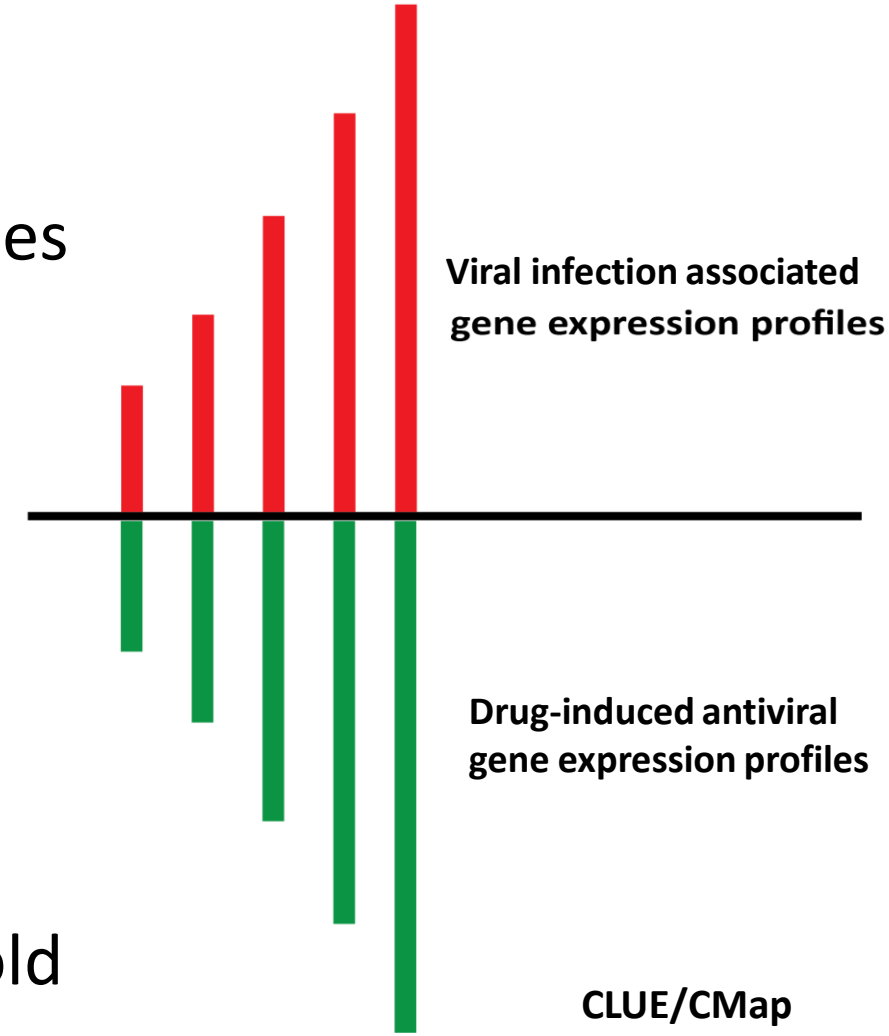
1. ***Understand what drug repurposing analyses are and how to use it as a tool for host-pathogen data***
2. ***Create a basic plot in R to visualize CLUE results***
3. ***Refine R plot components for specific colours, shapes and sizes***
4. ***Combine multiple plots in R and output as a file to the desktop***
5. ***Learn how to produce a similar figure in Tableau***

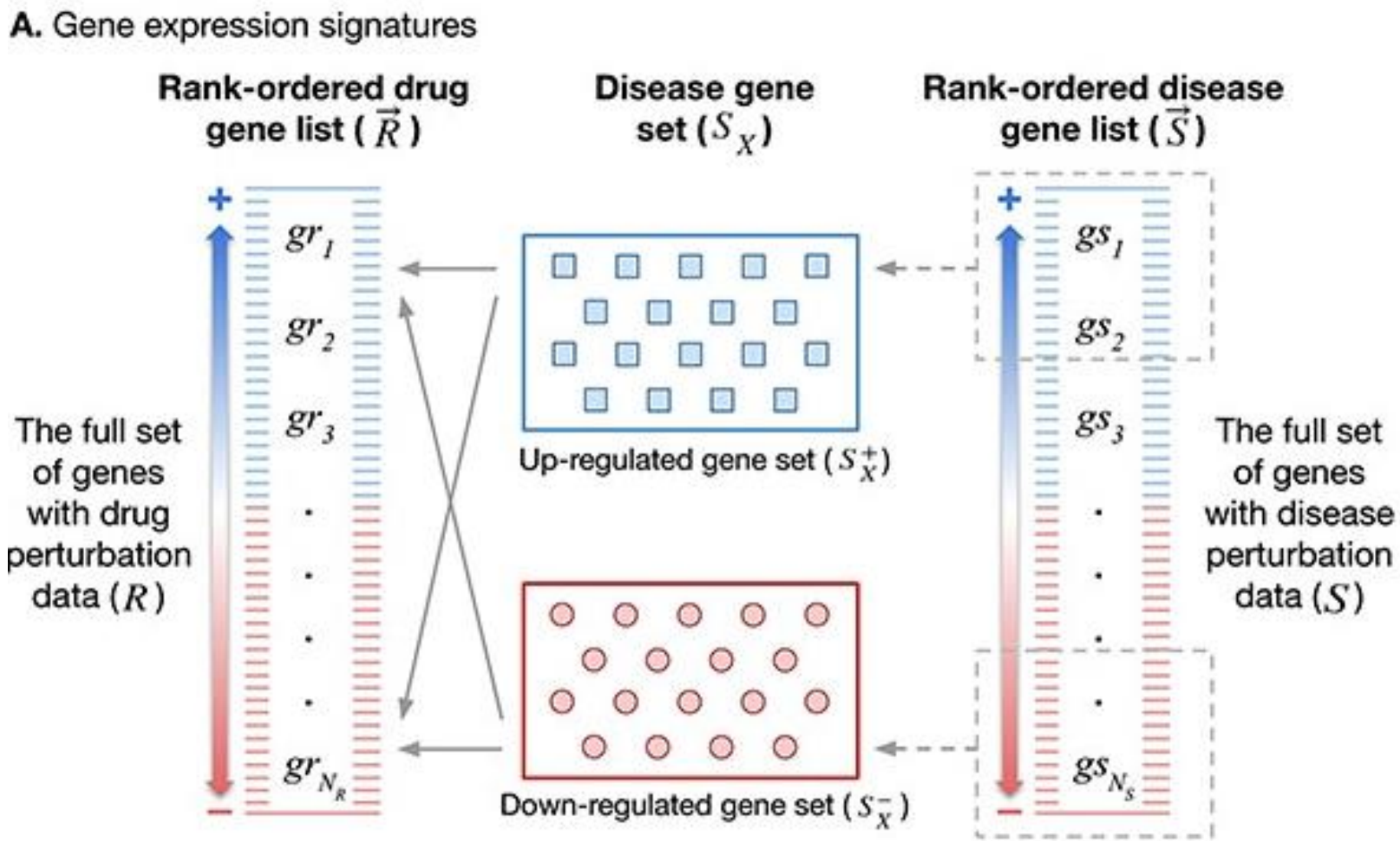
Drug Repurposing Analyses

Purpose: Identify potential small molecules that have the opposing gene expression profiles to the profile that you have used as input

Proof: inclusion of genetic evidence in drug development pipelines has been shown to increase success by at least two-fold

Gene expression: is determined by z-score, fold change, etc





Kewalin Samart, Phoebe Tuyishime, Arjun Krishnan, Janani Ravi, Reconciling multiple connectivity scores for drug repurposing, *Briefings in Bioinformatics*, Volume 22, Issue 6, November 2021, bbab161, https://doi.org/10.1093/bib/bbab161

frontiers in Immunology

Front Immunol. 2019; 10: 60.
Published online 2019 Jan 29. doi: 10.3389/fimmu.2019.00060
PMCID: PMC6361841
PMID: 30761132

Repurposing of Drugs as Novel Influenza Inhibitors From Clinical Gene Expression Infection Signatures

Andrés Pizzorno,[1,2,†] Olivier Terrier,[1,*†] Claire Nicolas de Lamballerie,[1,3] Thomas Julien,[1,4] Blandine Padey,[1,4] Aurélien Traversier,[1] Magali Roche,[3] Marie-Eve Hamelin,[2] Chantal Rhéaume,[2] Séverine Croze,[5] Vanessa Escuret,[1,6] Julien Poissy,[7] Bruno Lina,[1,6] Catherine Legras-Lachuer,[3,8] Julien Textoris,[9,10] and Manuel Rosa-Calatrava[1,4,*]

Review Article | Published: 04 March 2019

Drug repurposing for antimicrobial discovery

Maya A. Farha & Eric D. Brown

Nature Microbiology **4**, 565–577 (2019) | Cite this article

Research | Open Access | Published: 24 September 2021

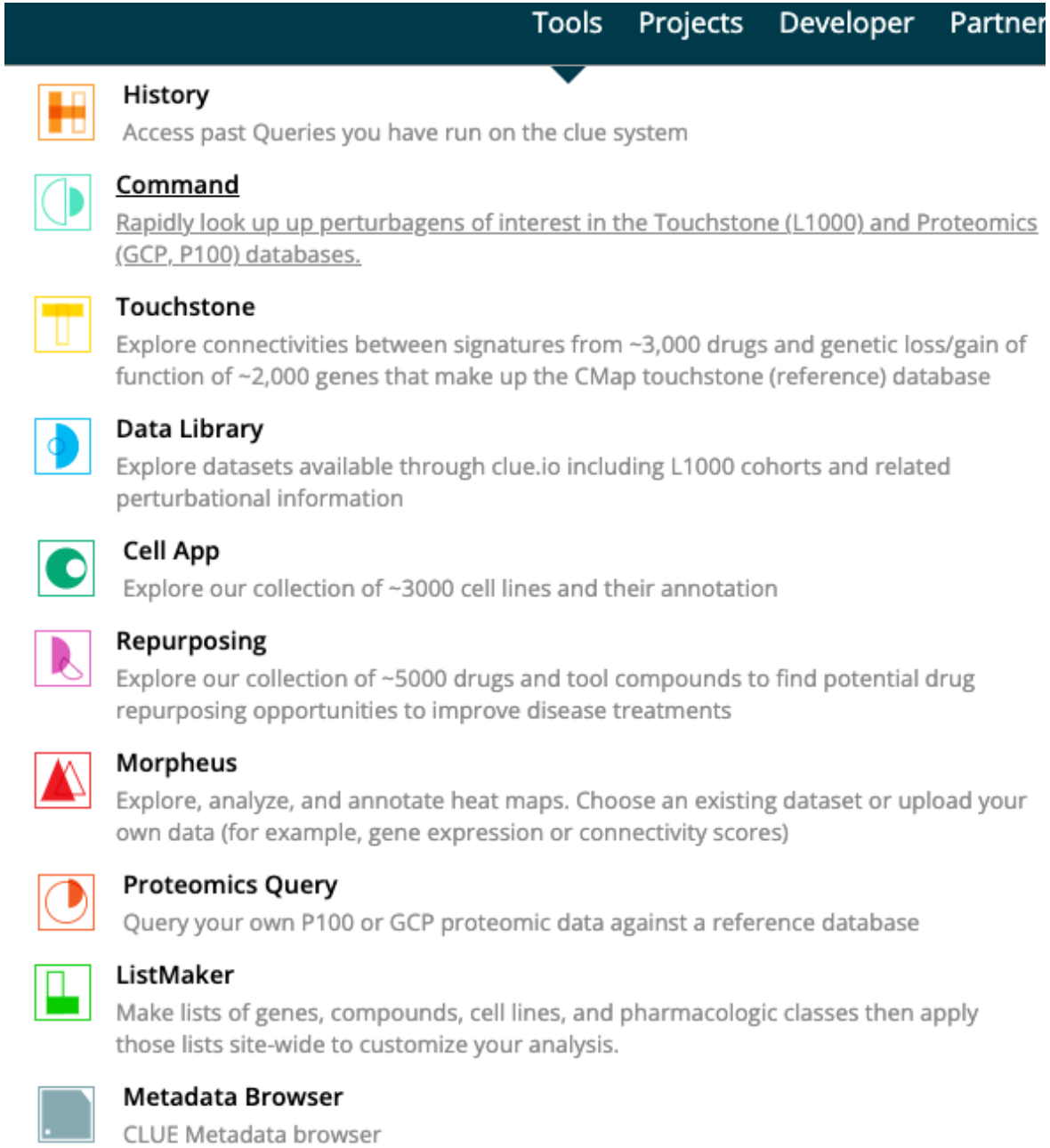
Clinical connectivity map for drug repurposing: using laboratory results to bridge drugs and diseases

Qianlong Wen, Ruoqi Liu & Ping Zhang

BMC Medical Informatics and Decision Making **21**, Article number: 263 (2021) | Cite this article

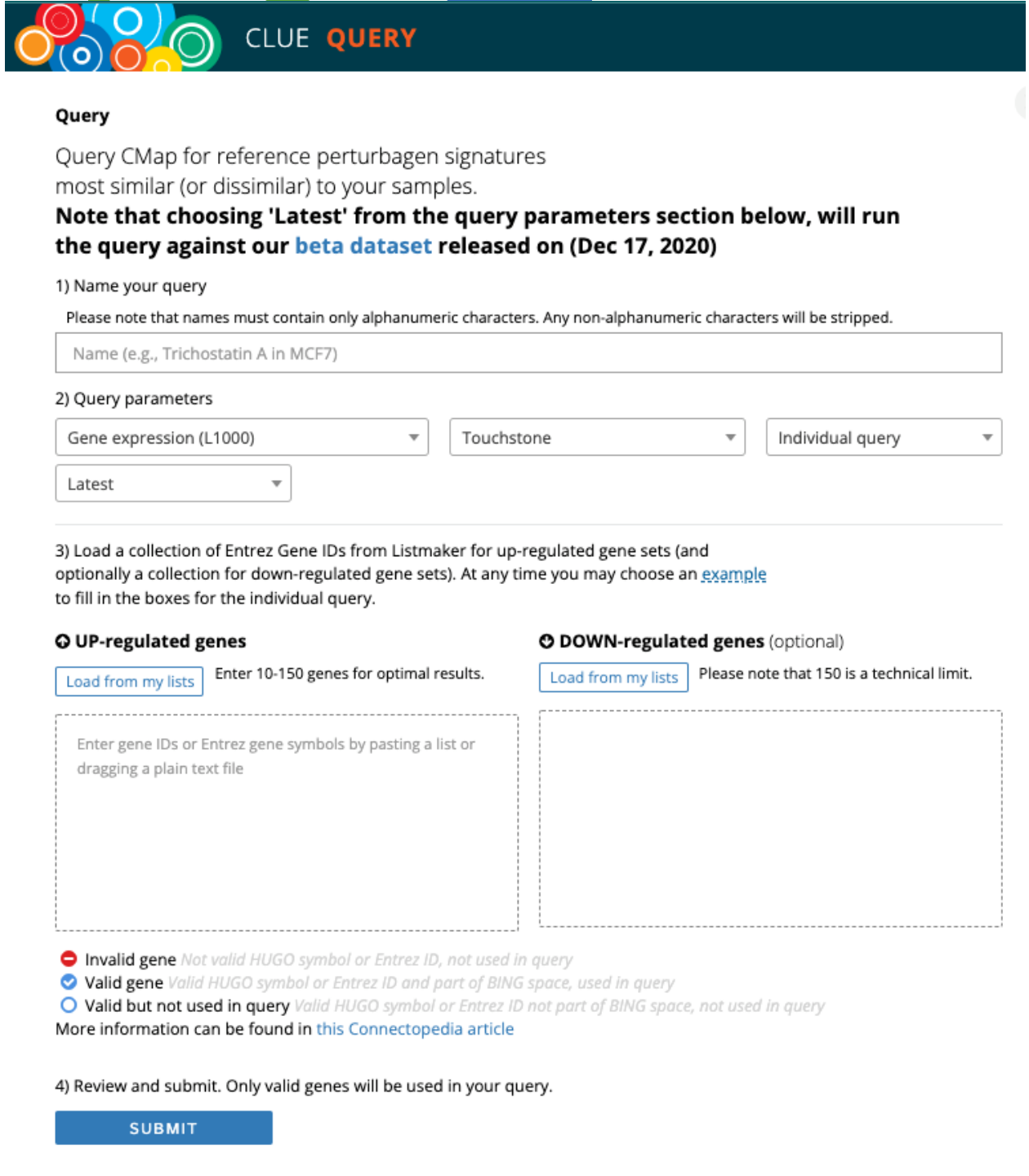
CLUE

- Cloud-based infrastructure
- Houses the CMap data
- Gene expression information for over 30,000 perturbagens



Four types of CMap Queries

1. Gene expression
   - Up and down regulated genes
2. Cell viability
   - Up and down regulated cell sets
3. Proteomics – Phosphorylation
   - GCT file
4. Proteomics – Histone Modifications
   - GCT file



PLOS ONE

PMID: 20957181

OPEN ACCESS PEER-REVIEWED

RESEARCH ARTICLE

Gene Expression Signature-Based Screening Identifies New Broadly Effective Influenza A Antivirals

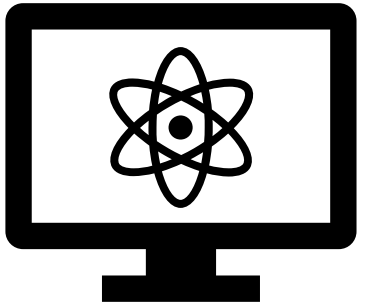
Laurence Josset, Julien Textoris, Béatrice Loriod, Olivier Ferraris, Vincent Moules, Bruno Lina, Catherine N'Guyen, Jean-Jacques Diaz, Manuel Rosa-Calatrava

Published: October 4, 2010 • https://doi.org/10.1371/journal.pone.0013169

Method: Used viral-infected human cells to obtain a list of differently expressed genes (via fold change)

Hypothesis: Perturbagens in CMap with an opposite gene expression to input may be candidate antivirals

Drug Repurposing Files to Visualize

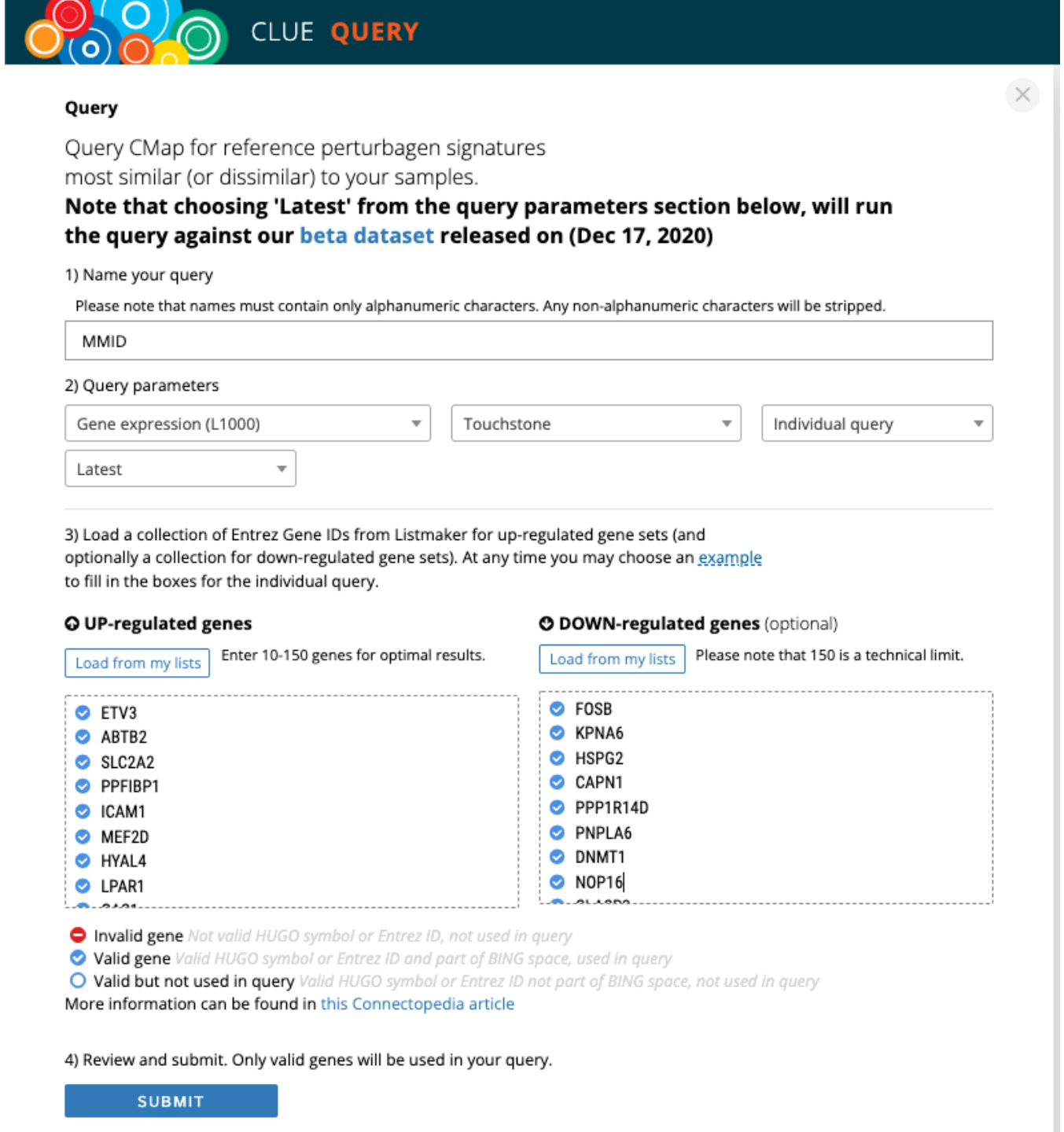


Josset *et al*., in supplementary table 5 have a list of up and down regulated genes that was input into the CLUE Query box

- As this experiment was done in a previous version of CLUE, two other down regulated genes had to be put in as requirements have changed for input

I am using an arbitrary cut off: **0.77**

This is just for learning purposes, not to replicate this study's results



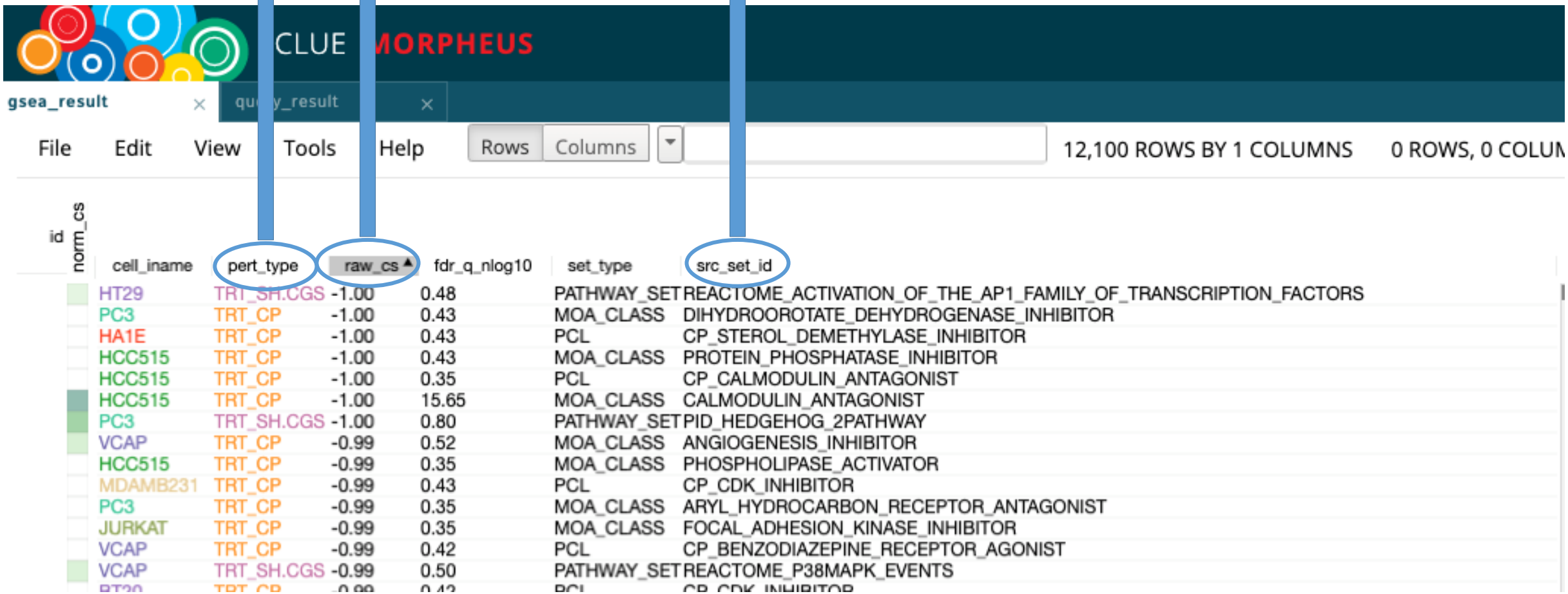
GSEA Results

Query Results

Perturbagen type

Median connectivity score

Mechanism of action



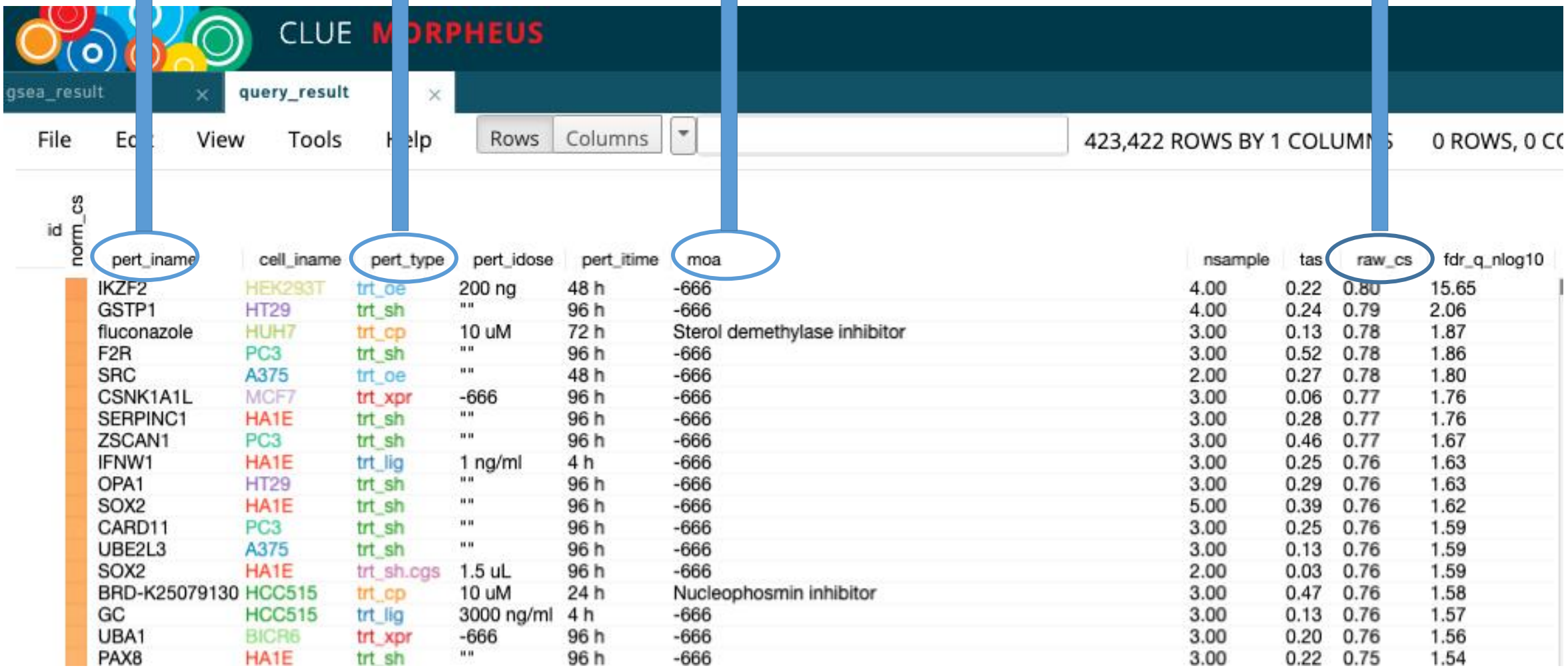
id	norm_cs	cell_iname	pert_type	raw_cs	fdr_q_nlog10	set_type	src_set_id
		HT29	TRT_SH.CGS	-1.00	0.48	PATHWAY_SET	REACTOME_ACTIVATION_OF_THE_AP1_FAMILY_OF_TRANSCRIPTION_FACTORS
		PC3	TRT_CP	-1.00	0.43	MOA_CLASS	DIHYDROOROTATE_DEHYDROGENASE_INHIBITOR
		HA1E	TRT_CP	-1.00	0.43	PCL	CP_STEROL_DEMETHYLASE_INHIBITOR
		HCC515	TRT_CP	-1.00	0.43	MOA_CLASS	PROTEIN_PHOSPHATASE_INHIBITOR
		HCC515	TRT_CP	-1.00	0.35	PCL	CP_CALMODULIN_ANTAGONIST
		HCC515	TRT_CP	-1.00	15.65	MOA_CLASS	CALMODULIN_ANTAGONIST
		PC3	TRT_SH.CGS	-1.00	0.80	PATHWAY_SET	PID_HEDGEHOG_2PATHWAY
		VCAP	TRT_CP	-0.99	0.52	MOA_CLASS	ANGIOGENESIS_INHIBITOR
		HCC515	TRT_CP	-0.99	0.35	MOA_CLASS	PHOSPHOLIPASE_ACTIVATOR
		MDAMB231	TRT_CP	-0.99	0.43	PCL	CP_CDK_INHIBITOR
		PC3	TRT_CP	-0.99	0.35	MOA_CLASS	ARYL_HYDROCARBON_RECEPTOR_ANTAGONIST
		JURKAT	TRT_CP	-0.99	0.35	MOA_CLASS	FOCAL_ADHESION_KINASE_INHIBITOR
		VCAP	TRT_CP	-0.99	0.42	PCL	CP_BENZODIAZEPINE_RECEPTOR_AGONIST
		VCAP	TRT_SH.CGS	-0.99	0.50	PATHWAY_SET	REACTOME_P38MAPK_EVENTS

Perturbagen

Perturbagen type

Mechanism of action

Median connectivity score

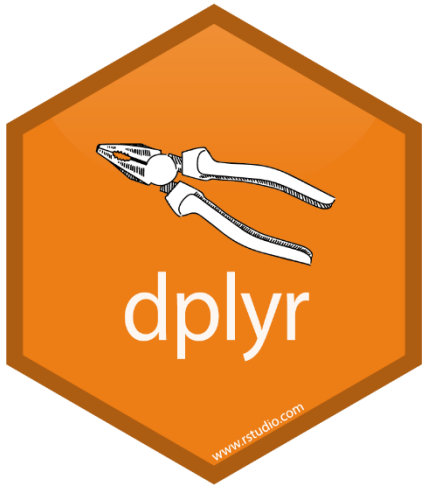


id	norm_cs	pert_iname	cell_iname	pert_type	pert_idose	pert_itime	moa	nsample	tas	raw_cs	fdr_q_nlog10
		IKZF2	HEK293T	trt_oe	200 ng	48 h	-666	4.00	0.22	0.80	15.65
		GSTP1	HT29	trt_sh	""	96 h	-666	4.00	0.24	0.79	2.06
		fluconazole	HUH7	trt_cp	10 uM	72 h	Sterol demethylase inhibitor	3.00	0.13	0.78	1.87
		F2R	PC3	trt_sh	""	96 h	-666	3.00	0.52	0.78	1.86
		SRC	A375	trt_oe	""	48 h	-666	2.00	0.27	0.78	1.80
		CSNK1A1L	MCF7	trt_xpr	-666	96 h	-666	3.00	0.06	0.77	1.76
		SERPINC1	HA1E	trt_sh	""	96 h	-666	3.00	0.28	0.77	1.76
		ZSCAN1	PC3	trt_sh	""	96 h	-666	3.00	0.46	0.77	1.67
		IFNW1	HA1E	trt_lig	1 ng/ml	4 h	-666	3.00	0.25	0.76	1.63
		OPA1	HT29	trt_sh	""	96 h	-666	3.00	0.29	0.76	1.63
		SOX2	HA1E	trt_sh	""	96 h	-666	5.00	0.39	0.76	1.62
		CARD11	PC3	trt_sh	""	96 h	-666	3.00	0.25	0.76	1.59
		UBE2L3	A375	trt_sh	""	96 h	-666	3.00	0.13	0.76	1.59
		SOX2	HA1E	trt_sh.cgs	1.5 uL	96 h	-666	2.00	0.03	0.76	1.59
		BRD-K25079130	HCC515	trt_cp	10 uM	24 h	Nucleophosmin inhibitor	3.00	0.47	0.76	1.58
		GC	HCC515	trt_lig	3000 ng/ml	4 h	-666	3.00	0.13	0.76	1.57
		UBA1	BICR6	trt_xpr	-666	96 h	-666	3.00	0.20	0.76	1.56
		PAX8	HA1E	trt_sh	""	96 h	-666	3.00	0.22	0.75	1.54

LEARNING OBJECTIVES

1. *Understand what drug repurposing analyses are and how to use it as a tool for host-pathogen data*
2. ***Create a basic plot in R to visualize CLUE results***
3. *Refine R plot components for specific colours, shapes and sizes*
4. *Combine multiple plots in R and output as a file to the desktop*
5. *Learn how to produce a similar figure in Tableau*

Explanation of Libraries



dplyr: A Grammar of Data Manipulation

- Core package of the tidyverse

snakecase: Convert Strings into any Case

- A string is an array of 1 or more characters (ex: words)

ggplot2: Main data visualization package in R

plotly: Helps make interactive figures/graphs in R

cowplot: Add on to ggplot2

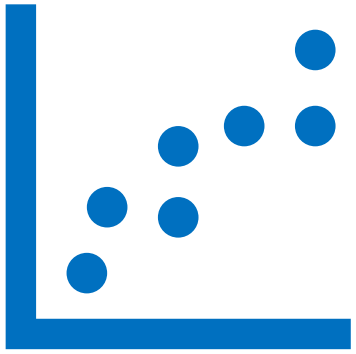
- To make figures publication ready
- Combine different types of plots

readxl: Reads in Excel files (.xls, .xlsx)

https://englelab.gatech.edu/useRguide/data-manipulation-using-dplyr.html; https://ggplot2.tidyverse.org/

Purpose of making a scatterplot

- To show different types of perturbagens
- To show median connectivity score range



- To show in blue/red colours, which perturbagens show a highly dissimilar expression profile and a highly similar expression profile

Going over to RStudio…

Loading in the Libraries

```
#Load in the libraries
library(dplyr) #install.packages("dyplr")
library(readxl) #install.packages("readxl")
library(snakecase) #install.packages("snakecase")
library(ggplot2) #install.packages("ggplot2")
library(cowplot) #install.packages("cowplot")
library(plotly) #install.packages("plotly")
```

Make a scatterplot of sheet 1 results

Step 1. Read in Excel file and see what data looks like

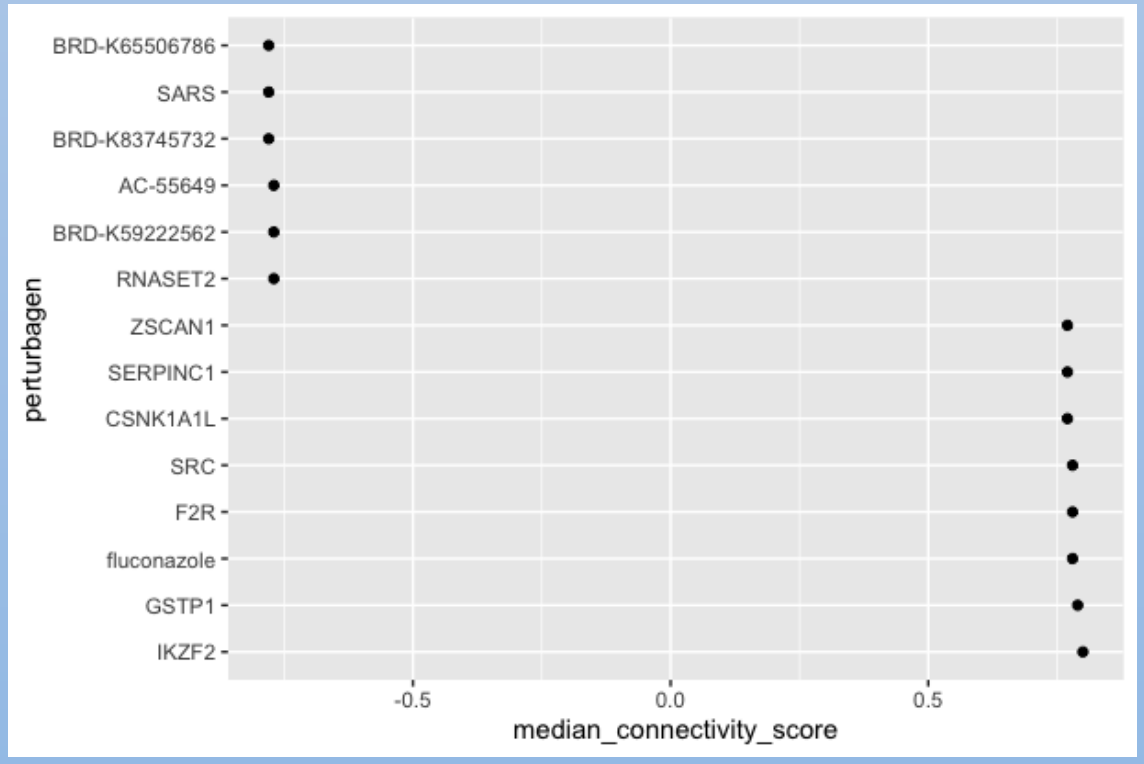
```
#Step 1. Read in Excel File
clue_results<-read_excel("/Users/msarvis/Desktop/MMID/mmid_workshop.xls", sheet=1) %>%
  rename_with(to_snake_case)

#See what it looks like
glimpse(clue_results)
```

Step 2. Change the perturbagen names into a factor and give it levels

```
clue_results$perturbagen <- factor(clue_results$perturbagen, levels = clue_results$perturbagen)
```

Step 3. Create a basic scatterplot with ggplot2

```
#Step 2. Create a basic plot
basic_plot<-ggplot(clue_results, aes(x=median_connectivity_score, y=perturbagen)) +
  geom_point()

#View the plot
basic_plot
```



LEARNING OBJECTIVES

1. *Understand what drug repurposing analyses are and how to use it as a tool for host-pathogen data*
2. *Create a basic plot in R to visualize CLUE results*
3. ***Refine R plot components for specific colours, shapes and sizes***
4. *Combine multiple plots in R and output as a file to the desktop*
5. *Learn how to produce a similar figure in Tableau*

Step 4. Change the colour of the perturbagens based on their connectivity score

```
colours <- ifelse(clue_results$median_connectivity_score < 0, "blue", "red")
```

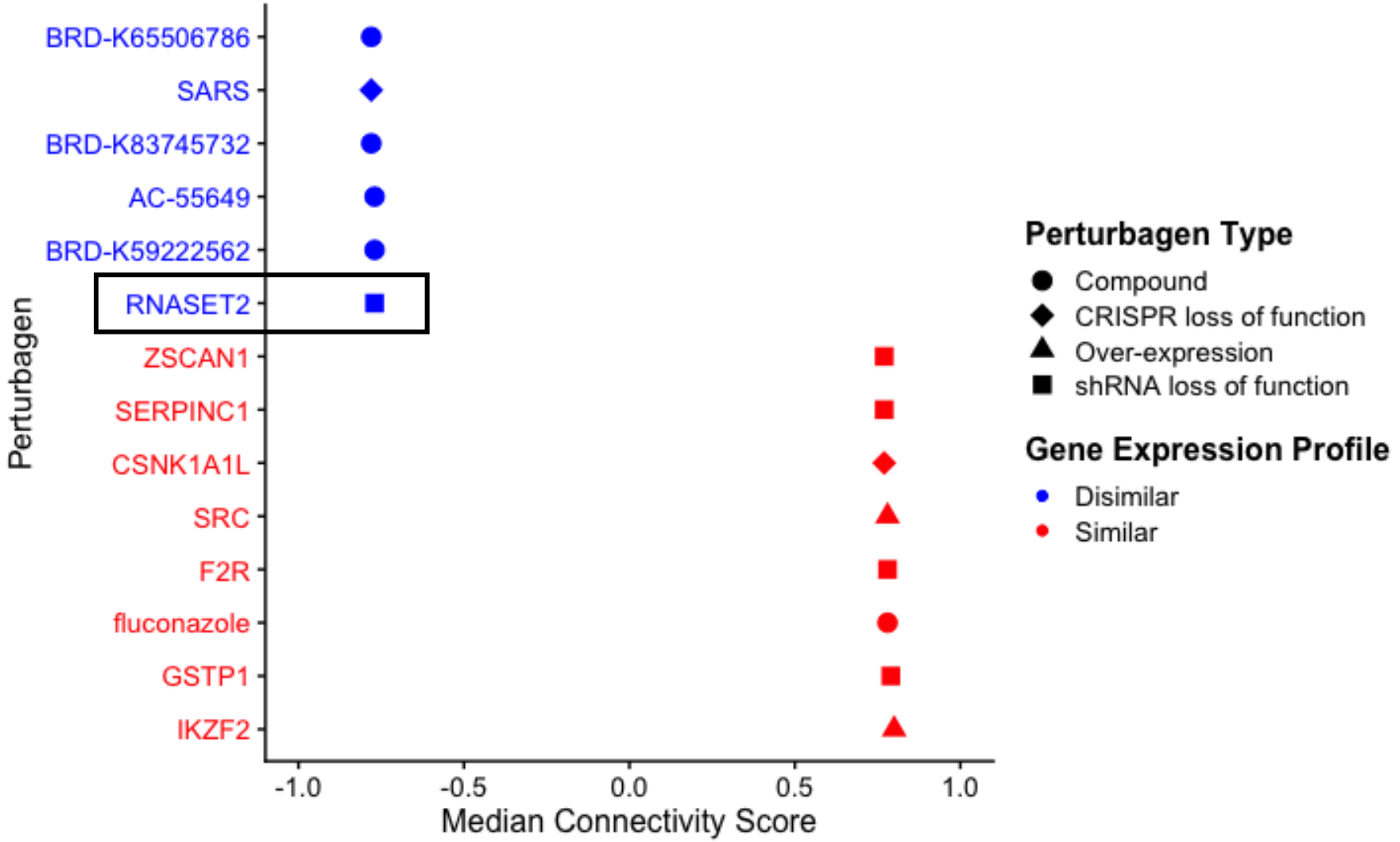
Step 5. Adding in details for colour, shape and sizes

```
drug_repurposing_plot<-basic_plot + geom_point(aes(colour=gene_expression_profile, shape= perturbagen_type, size= perturbagen_type)) +
  theme_cowplot(12) + ggtitle("Top Antiviral Drug Repurposing Results")+
    xlab("Median Connectivity Score") +ylab ("Perturbagen") + xlim(-1,1) +
      theme(plot.title = element_text(size=20, hjust = 0.5), axis.text.y = element_text(hjust = 1, colour=colours),
            legend.title = element_text(face = "bold")) +
               scale_shape_manual(values= c(19,18,17,15))+
                  scale_color_manual(values = c("Disimilar" = "blue", "Similar" = "red")) +
                      scale_size_manual(values=c(3,4,3,3)) + guides(shape=guide_legend("Perturbagen Type")) +
                        guides(size=guide_legend("Perturbagen Type")) +
                         guides(color=guide_legend("Gene Expression Profile"))
```

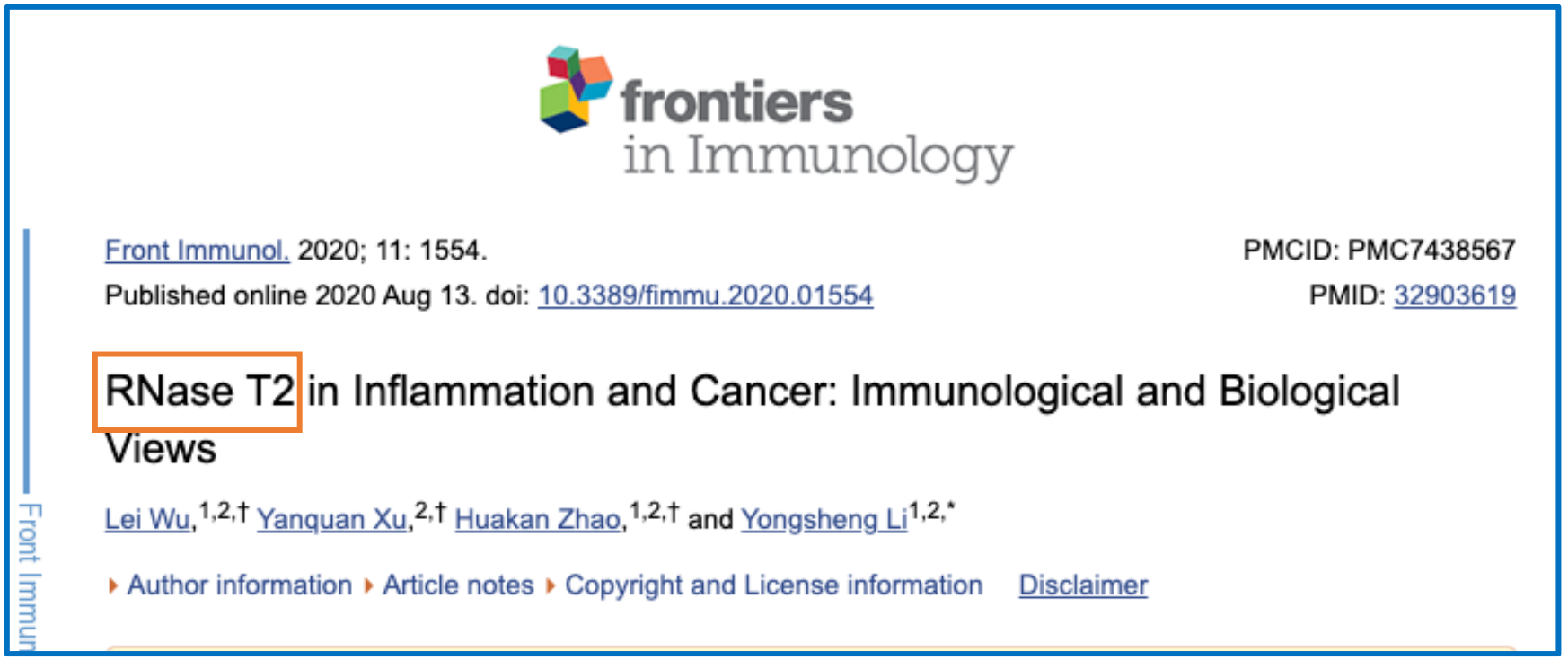
```
ggplotly(drug_repurposing_plot)
```

*Bonus: See it interactively!

Top Antiviral Drug Repurposing Results



What can we learn from the scatterplot?



frontiers in Immunology

Front Immunol. 2020; 11: 1554.
Published online 2020 Aug 13. doi: 10.3389/fimmu.2020.01554
PMCID: PMC7438567
PMID: 32903619

RNase T2 in Inflammation and Cancer: Immunological and Biological Views

Lei Wu,[1,2,†] Yanquan Xu,[2,†] Huakan Zhao,[1,2,†] and Yongsheng Li[1,2,*]

Author information · Article notes · Copyright and License information Disclaimer

(10, 11). Following tissue damage, RNase T2 is secreted and participates in resistance against RNA viruses or functions as an alarm signaling molecule to regulate the host immune response and contributes to tissue remodeling and repair (12, 13).

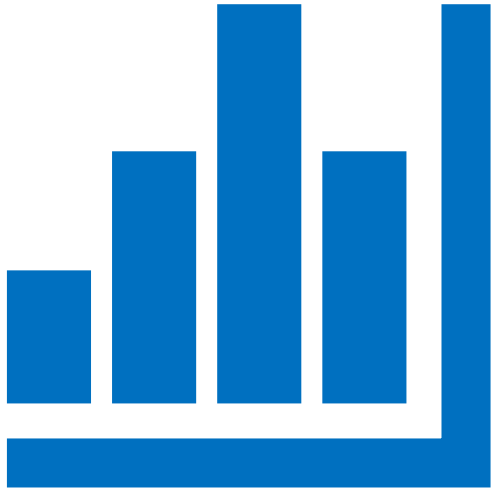
The antiviral activities of RNase T2 have also been reported (3, 62). Elevated levels of extracellular RNase T2 expression resulted in increased resistance to Cucumovirus and Virgaviridae infection in plants (62). In

LEARNING OBJECTIVES

1. *Understand what drug repurposing analyses are and how to use it as a tool for host-pathogen data*
2. *Create a basic plot in R to visualize CLUE results*
3. *Refine R plot components for specific colours, shapes and sizes*
4. ***Combine multiple plots in R and output as a file to the desktop***
5. *Learn how to produce a similar figure in Tableau*

Purpose of making a bar plot

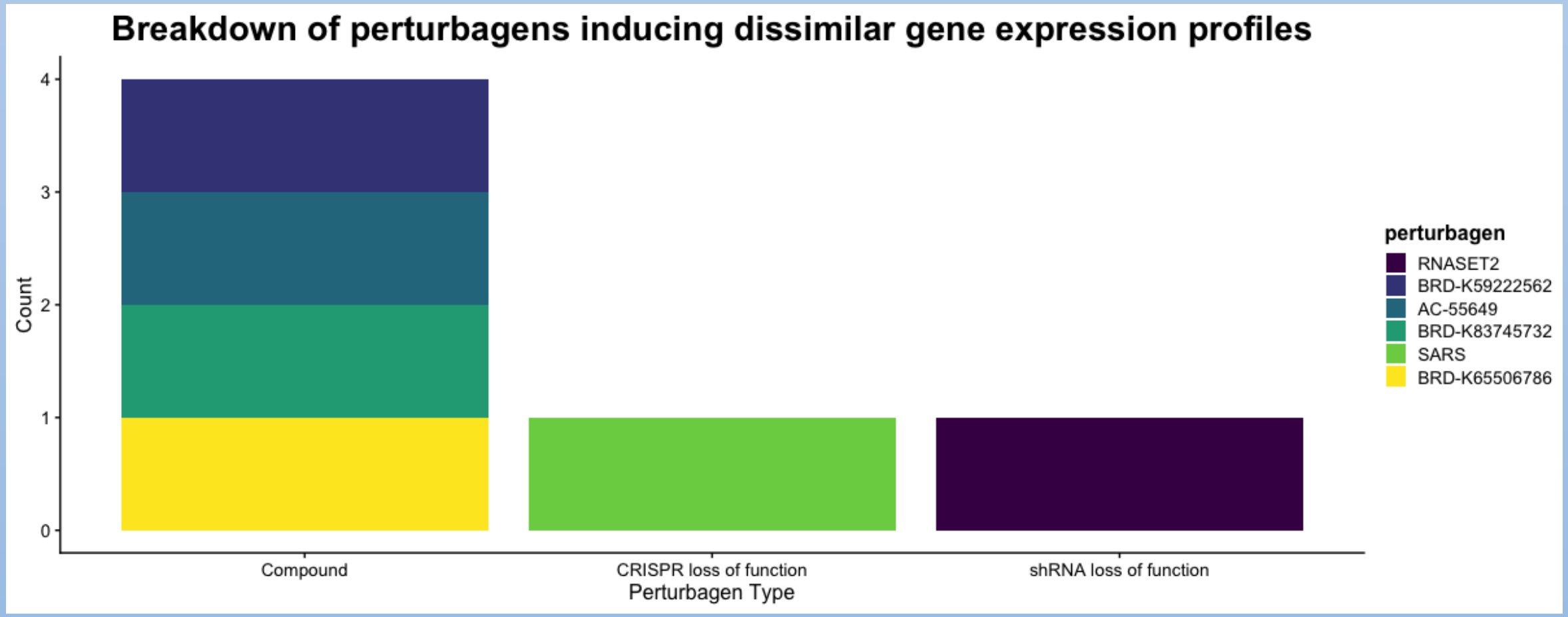
- Compare between similar and dissimilar perturbagen types
  - How many perturbagens are in each category
  - Which perturbagen types are the most/least common
  - How the perturbagen types differ between expression profiles

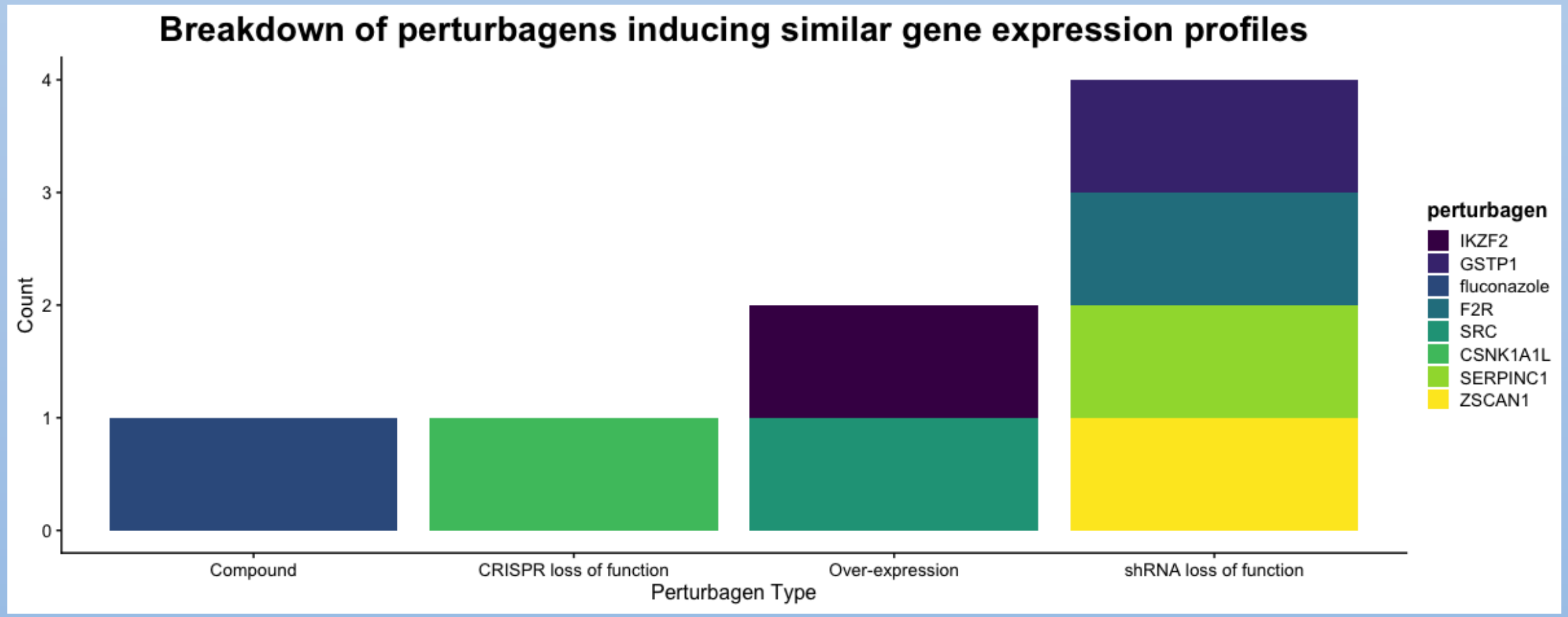


Going over to RStudio…

Step 6. Create a bar plot to compare the different gene expression profiles

```
#a. Filter results to perturbagens inducing dissimilar (and similar) gene expression profiles (i.e. blue and red perturbagens)
disimilar_perturbagens<-clue_results %>% filter(median_connectivity_score<0)
similar_perturbagens<-clue_results %>% filter(median_connectivity_score>0)

#b. Make a basic bar graph
basic_bar_graph<-ggplot(disimilar_perturbagens, aes(x=perturbagen_type, fill=perturbagen)) +
  geom_bar()

#c. Add in different elements you want to have
drug_repurposing_bar_plot_dissimilar<-basic_bar_graph + geom_bar() + theme_cowplot(12) +
                              ggtitle("Breakdown of perturbagens inducing dissimilar gene expression profiles") +
                                  xlab("Perturbagen Type") + ylab ("Count") +
                                    theme (plot.title = element_text(size=20,hjust = 0.5),
                                        legend.title = element_text(face = "bold")) +
                                            scale_fill_viridis_d()

#Bonus: See it interactively
ggplotly(drug_repurposing_bar_plot_dissimilar)
```


Breakdown of perturbagens inducing dissimilar gene expression profiles
Count
4
3
2
1
0
Compound
CRISPR loss of function
shRNA loss of function
Perturbagen Type
perturbagen
RNASET2
BRD-K59222562
AC-55649
BRD-K83745732
SARS
BRD-K65506786


Breakdown of perturbagens inducing similar gene expression profiles
Count
0
1
2
3
4
Compound
CRISPR loss of function
Over-expression
shRNA loss of function
Perturbagen Type
perturbagen
IKZF2
GSTP1
fluconazole
F2R
SRC
CSNK1A1L
SERPINC1
ZSCAN1

What can we learn from the bar plot?

- Dissimilar gene expression profile perturbagens
  - Most common: compound
  - Less perturbagens (n=6)
  - 3 types of perturbagens

- Similar gene expression profile perturbagens
  - Most common: shRNA loss of function
  - More perturbagens (n=8)
  - 4 types of perturbagens

Step 7. Combine the plots together and save to your computer desktop

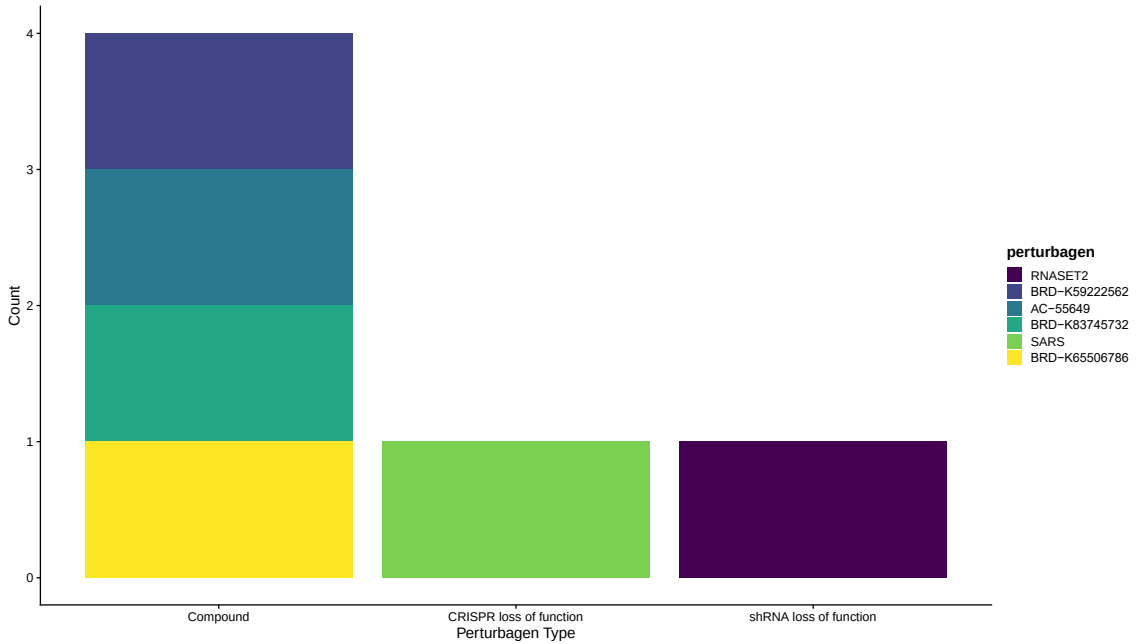
```
bottom_row <- plot_grid(drug_repurposing_bar_plot_dissimilar, drug_repurposing_bar_plot_similar,
                        labels = c('2', '3'), label_size = 12)
plot_grid(drug_repurposing_plot, bottom_row,
          labels = c('1', ''), label_size=12, ncol = 1)

ggsave("/Users/msarvis/Desktop/MMID/drug_repurposing.pdf", height=15, width=25, units='in', dpi=300)
```

1

Top Antiviral Drug Repurposing Results

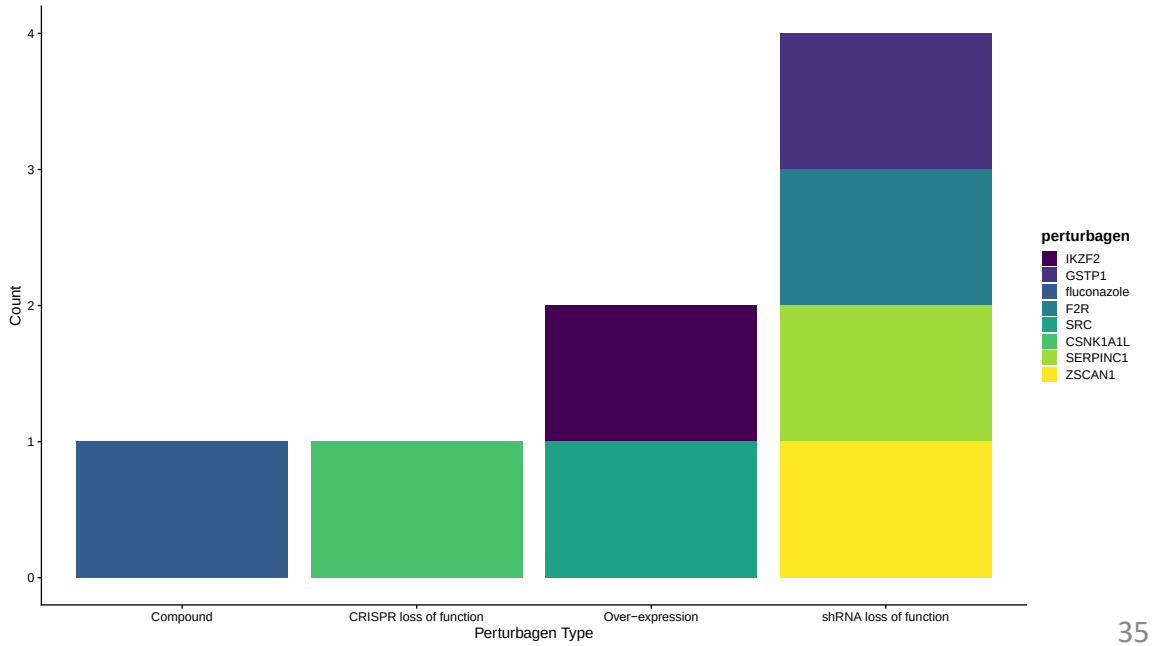
2

Breakdown of perturbagens inducing dissimilar gene expression profiles



3

Breakdown of perturbagens inducing similar gene expression profiles



LEARNING OBJECTIVES

1. *Understand what drug repurposing analyses are and how to use it as a tool for host-pathogen data*
2. *Create a basic plot in R to visualize CLUE results*
3. *Refine R plot components for specific colours, shapes and sizes*
4. *Combine multiple plots in R and output as a file to the desktop*
5. ***Learn how to produce a similar figure in Tableau***

RStudio (ggplot2) Vs Tableau

- More control
- Can save to desktop
- Can do data analysis
- Can be interactive with Plotly or Shiny

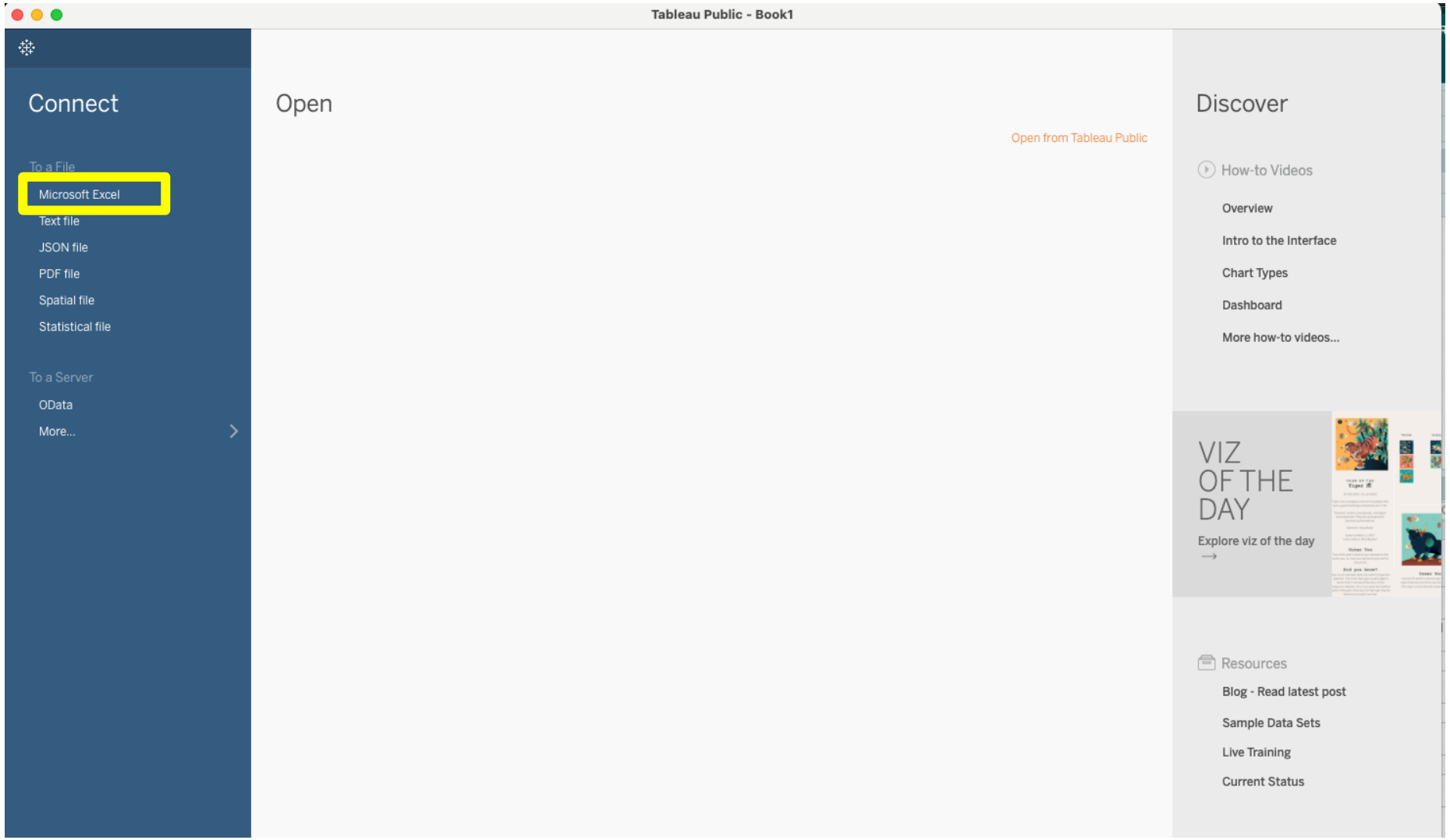
- Need to code

- Artistic freedom
- Dashboard is interactive
- Less coding

- Less control
- Only saves to online website
- Limited capabilities with free version

Going over to Tableau…

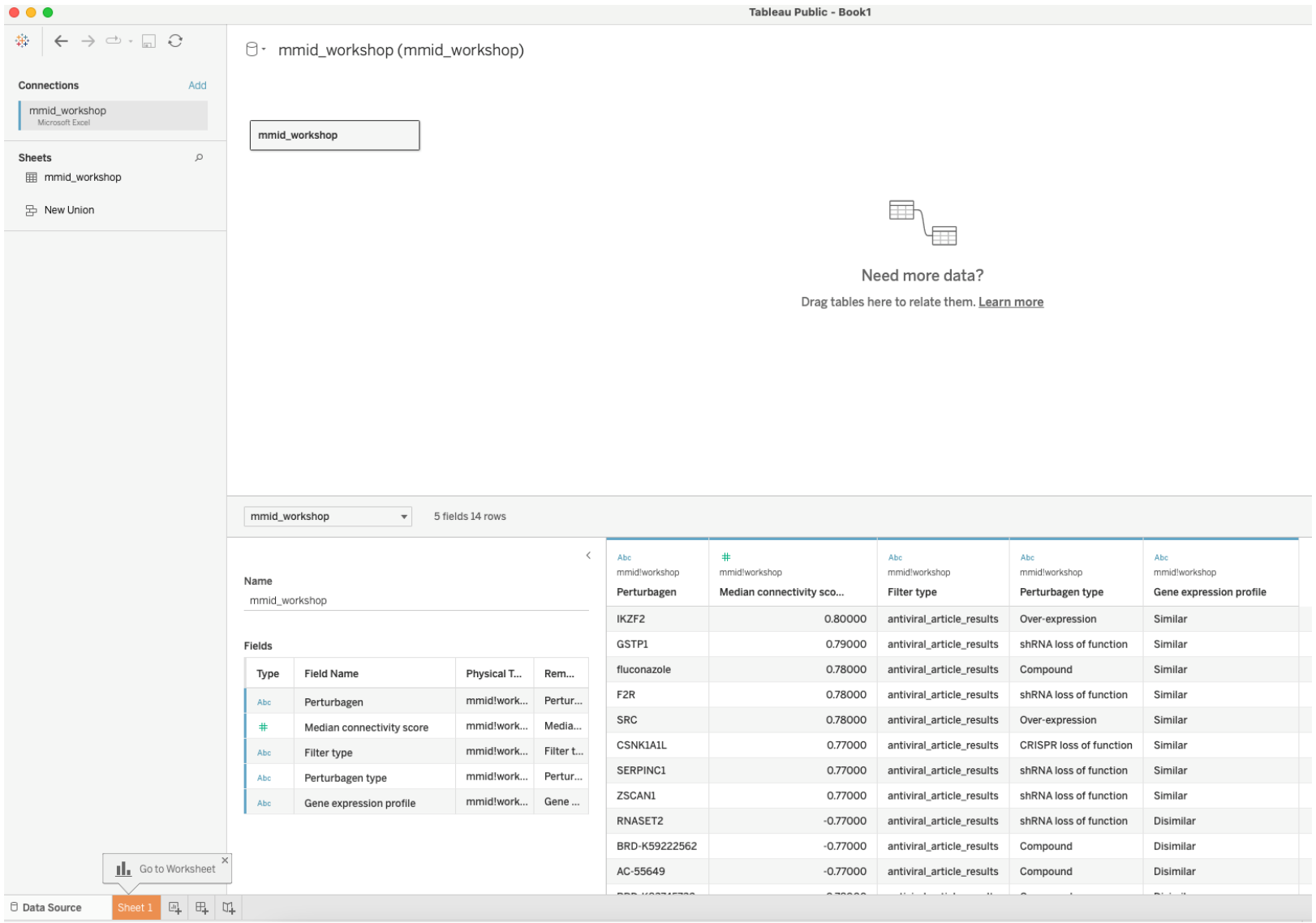
1. Import your Excel file

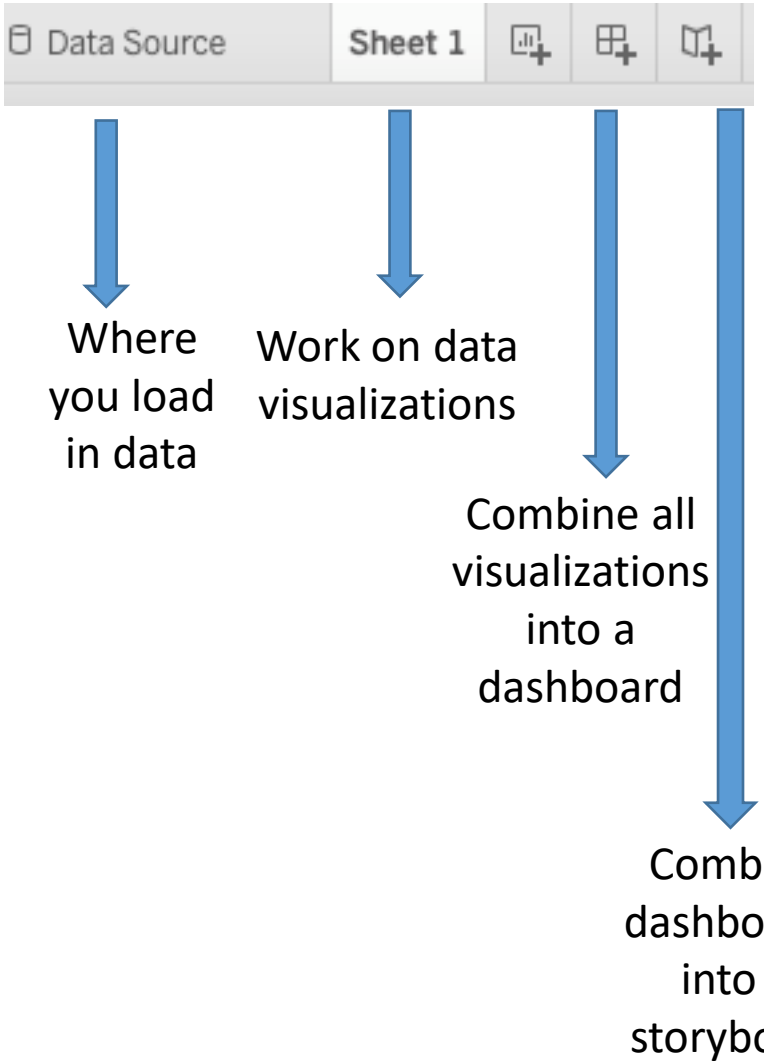


https://public.tableau.com/en-us/s/

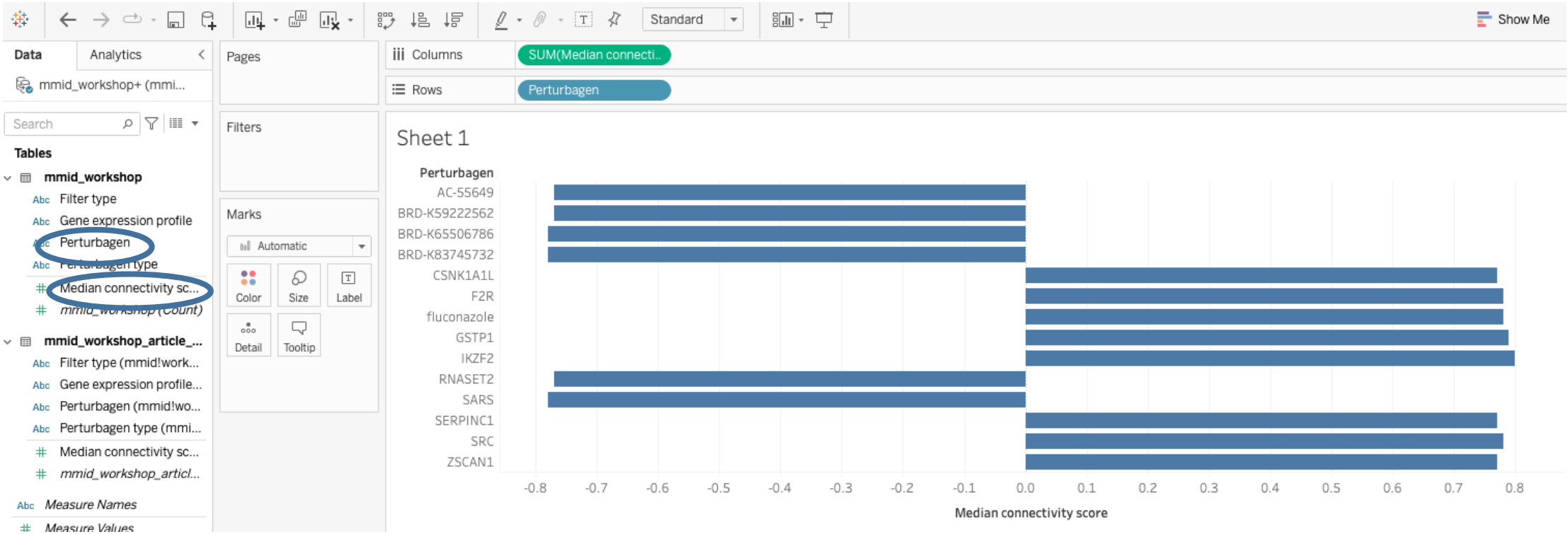
Enter an email and the app will install! Also make an account to save visualizations.

2. Put in your sheet and look at potential options

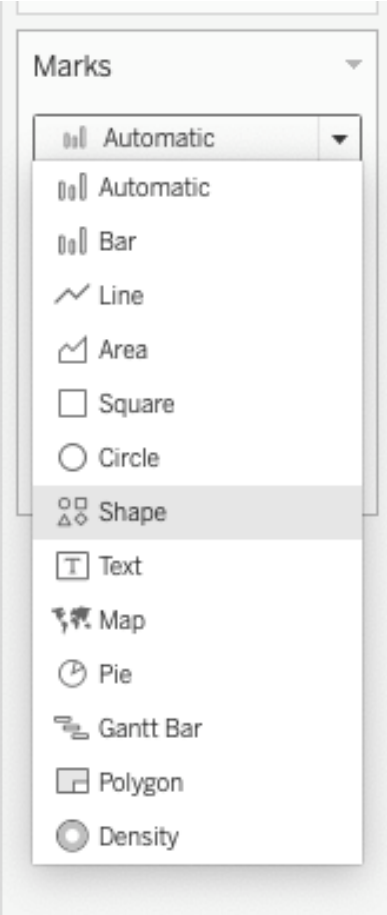




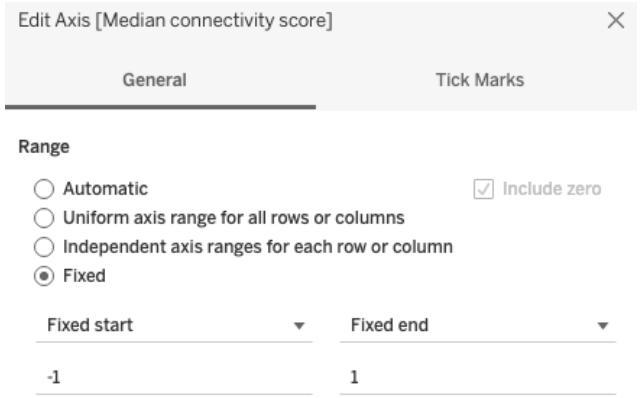
3. Create a scatterplot of results in Worksheet 1



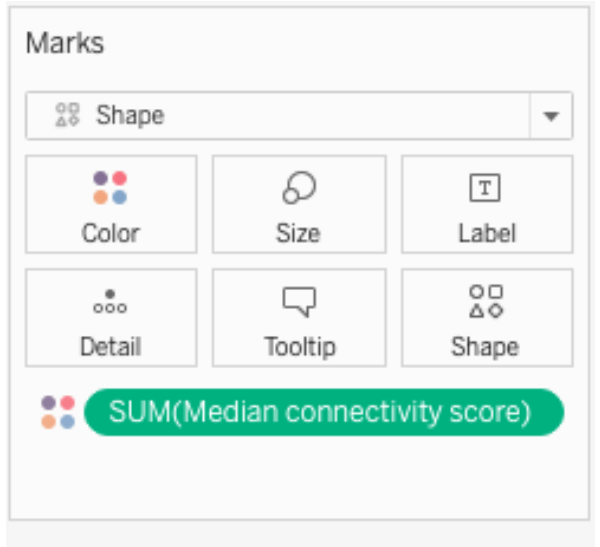
Add in colours, shapes, etc



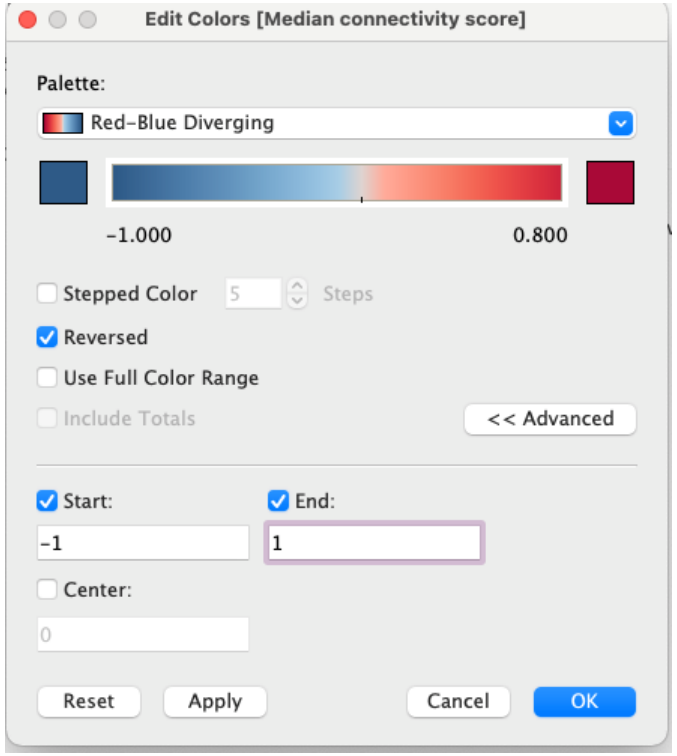
a. Choose graph type



b. Make x axis limits and rename axis



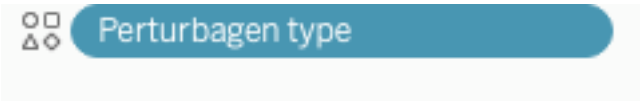
c. Make colours based on median connectivity score



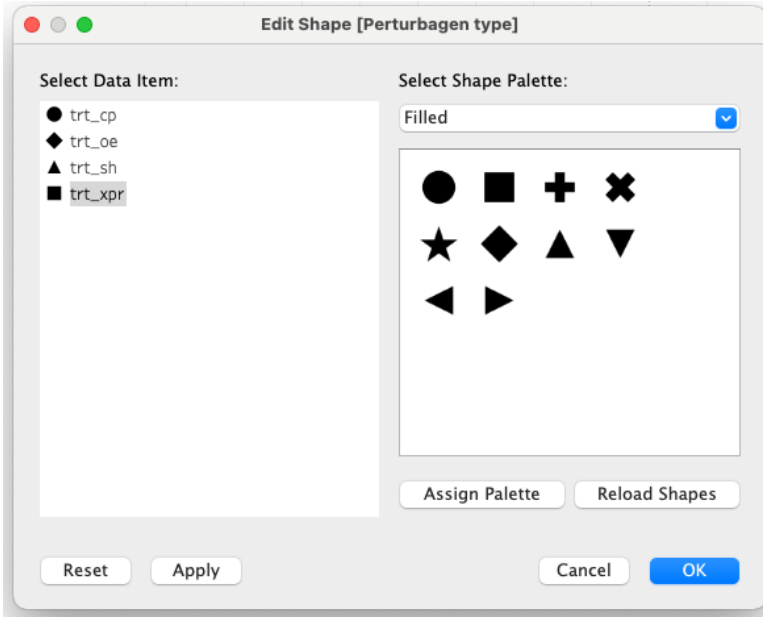
d. Make negative colours blue and positive colours red

Add in colours, shapes, etc

e. Sort so that the connectivity scores are ordered as you like

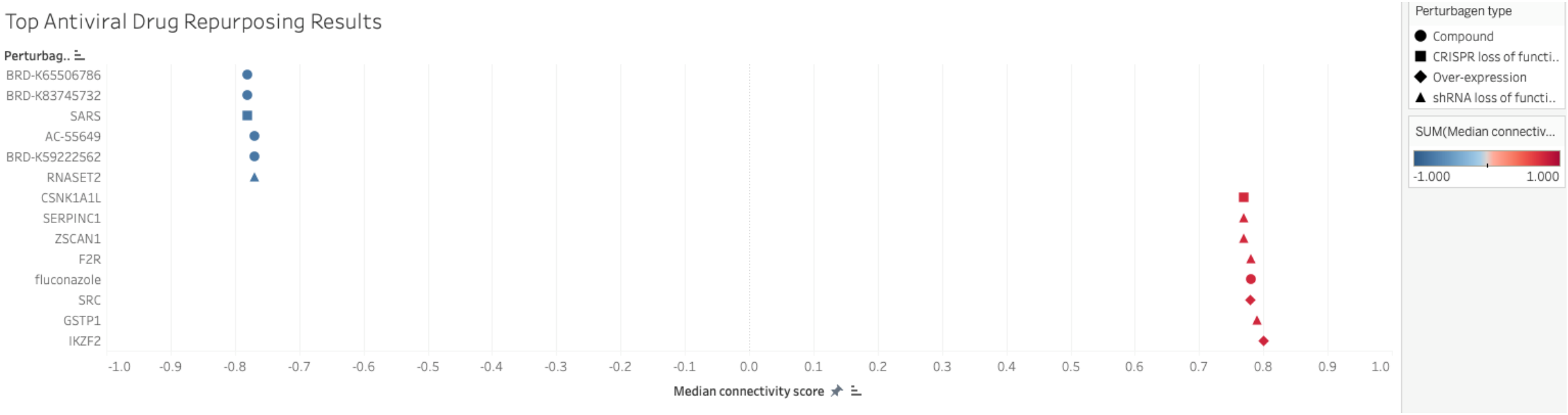


f. Make perturbagen type a specific shape

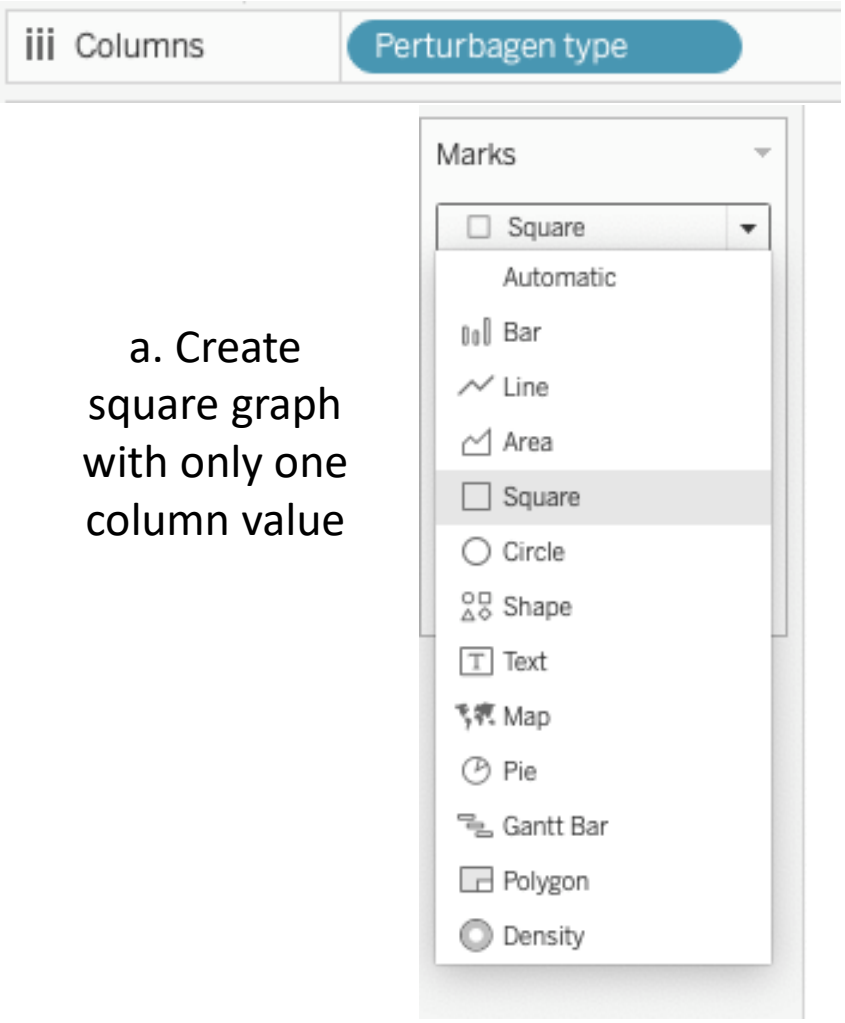


g. Select specific shapes for each perturbagen type

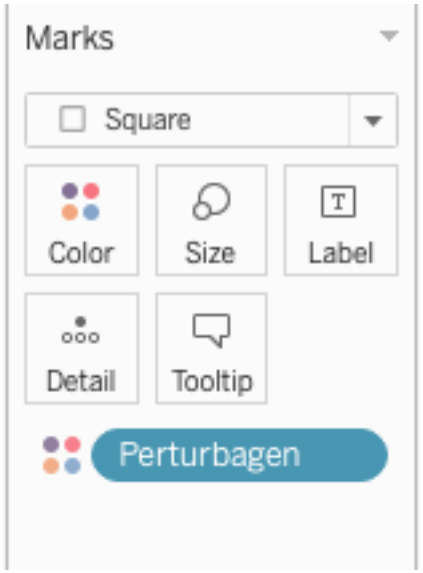
Final Scatterplot!



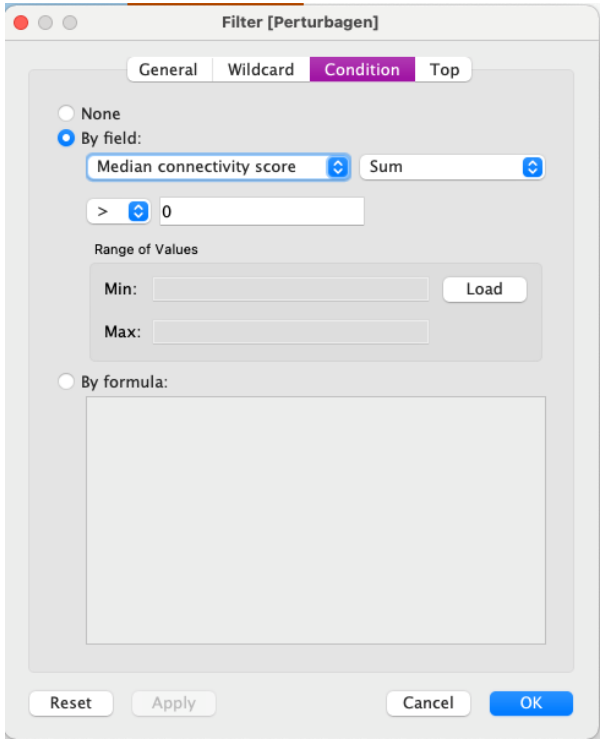
4. Create bar charts of data



a. Create square graph with only one column value

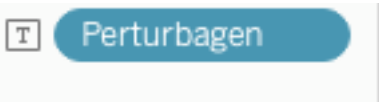


b. Make perturbagen a square colour



c. Filter perturbagens in the bar graph by connectivity score

d. Increase the square sizes

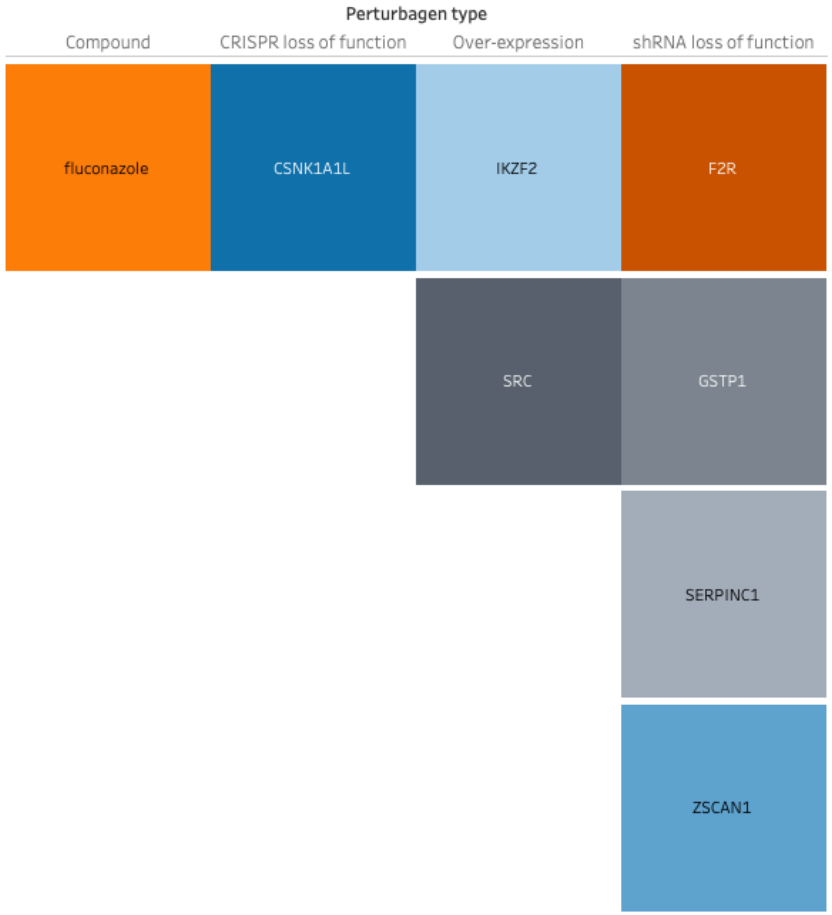


e. Write the perturbagen names on the graph

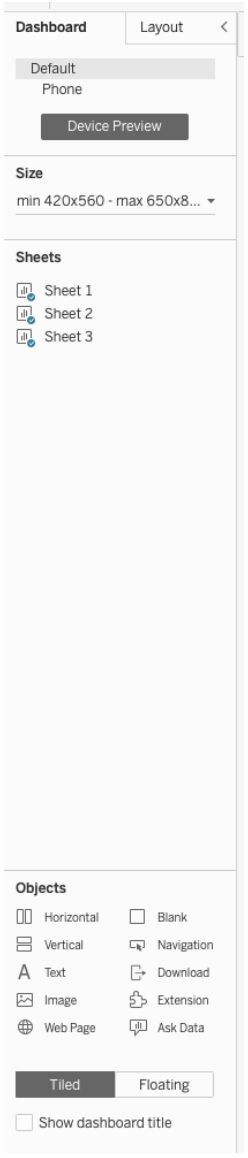
Bar charts produced in Tableau

Breakdown of perturbagens inducing dissimilar gene expression profiles

Breakdown of perturbagens inducing similar gene expression profiles

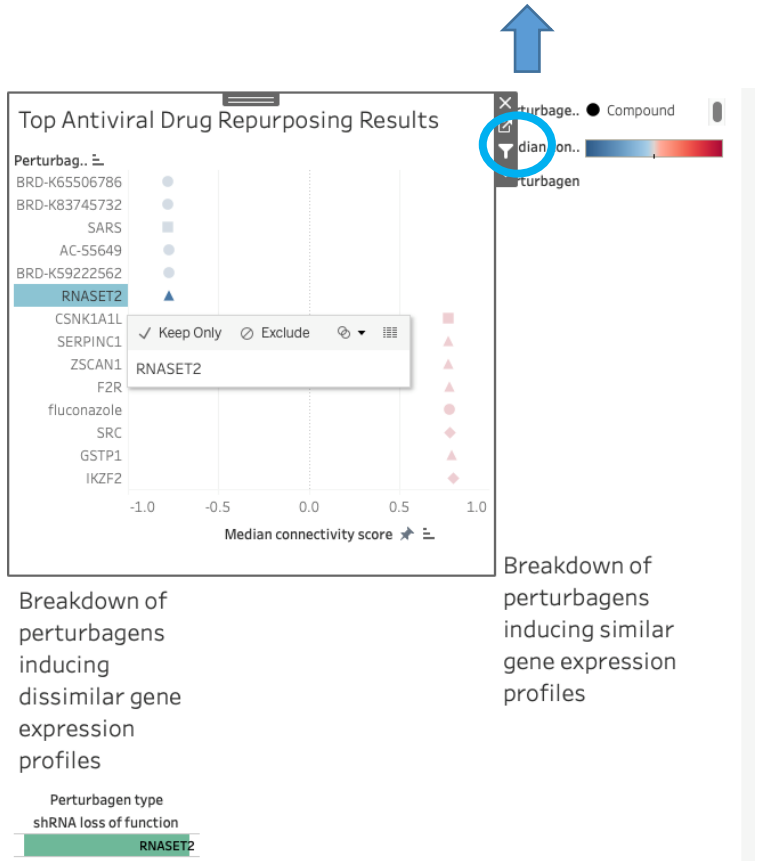


5. Combine all graphs on Dashboard



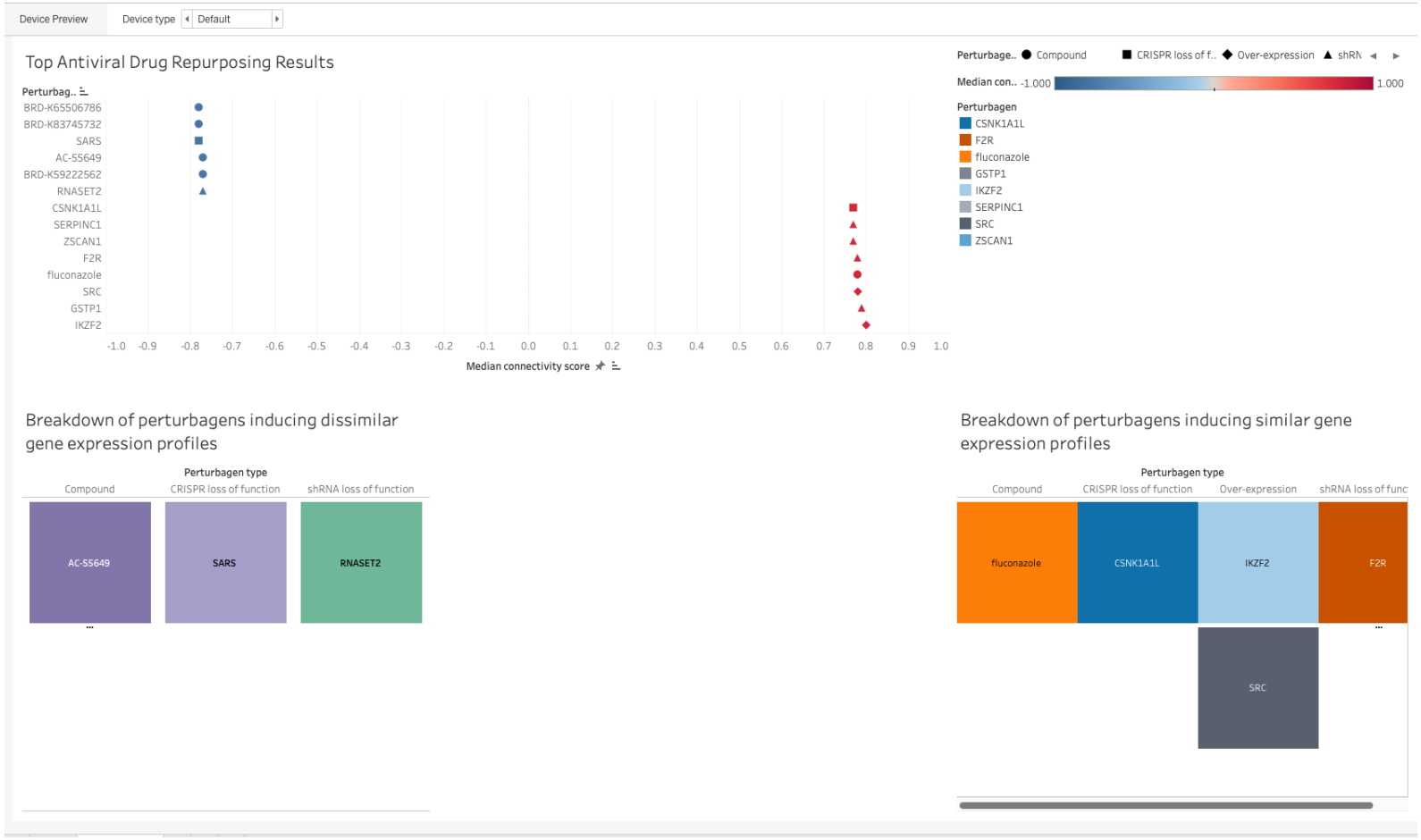
a. Different layouts you can have as well as objects you can insert (i.e. hyperlink, extension, etc)

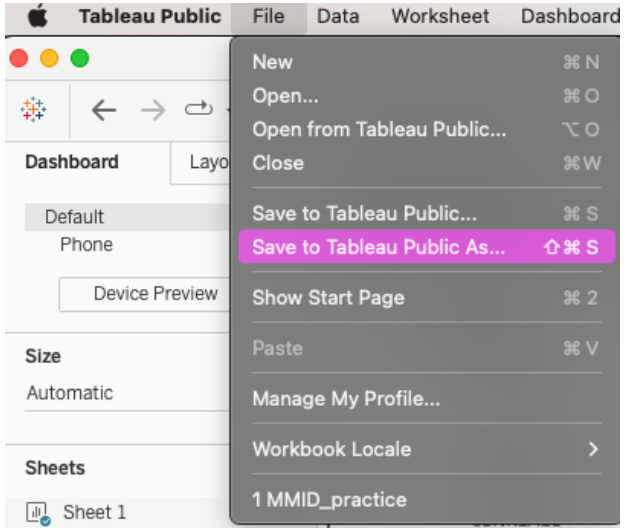
Interactive element!



b. Clicking the filter button on one graph and clicking on a perturbagen, can bring that one up on another graph

5. Combine all graphs on Dashboard





c. Save dashboard to Tableau public account

Tableau vs RStudio Drug Repurposing Graphs

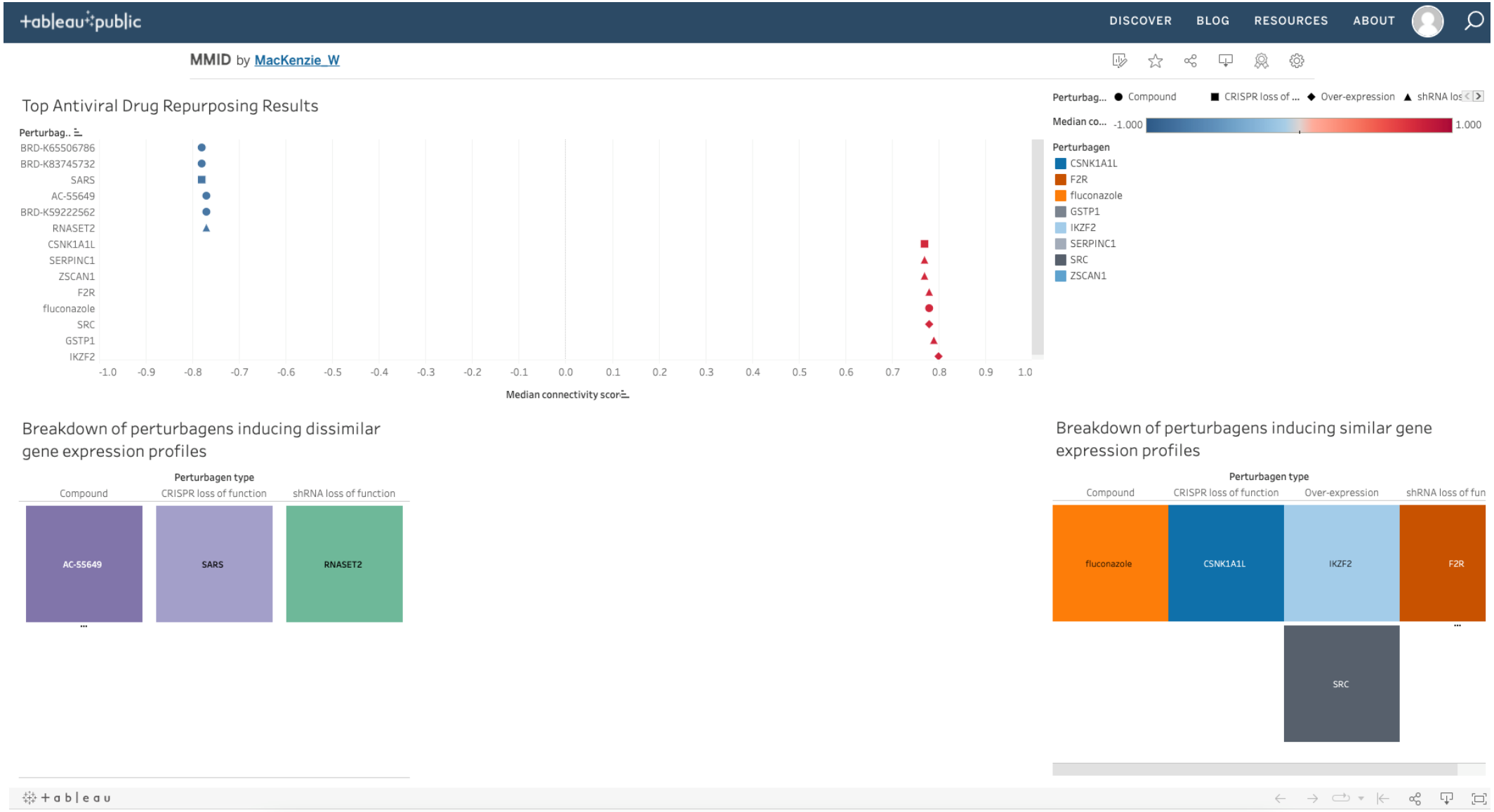
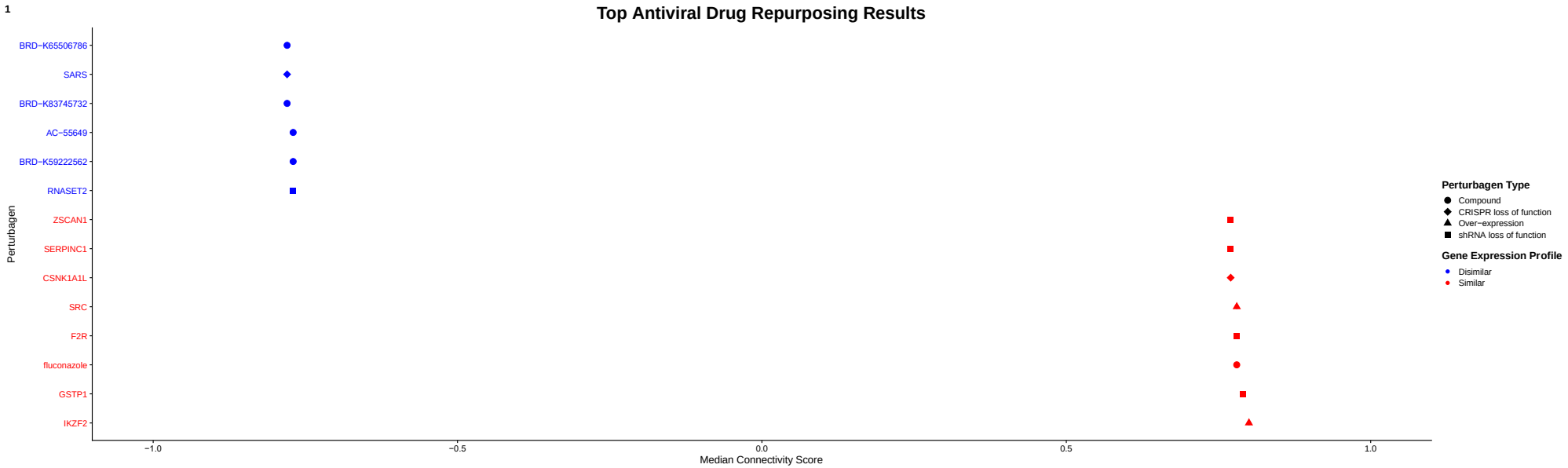
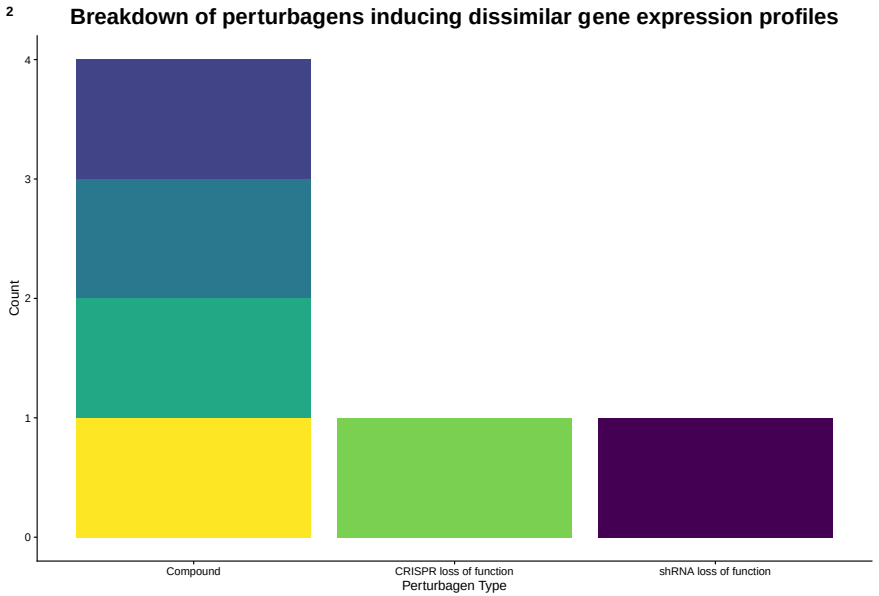
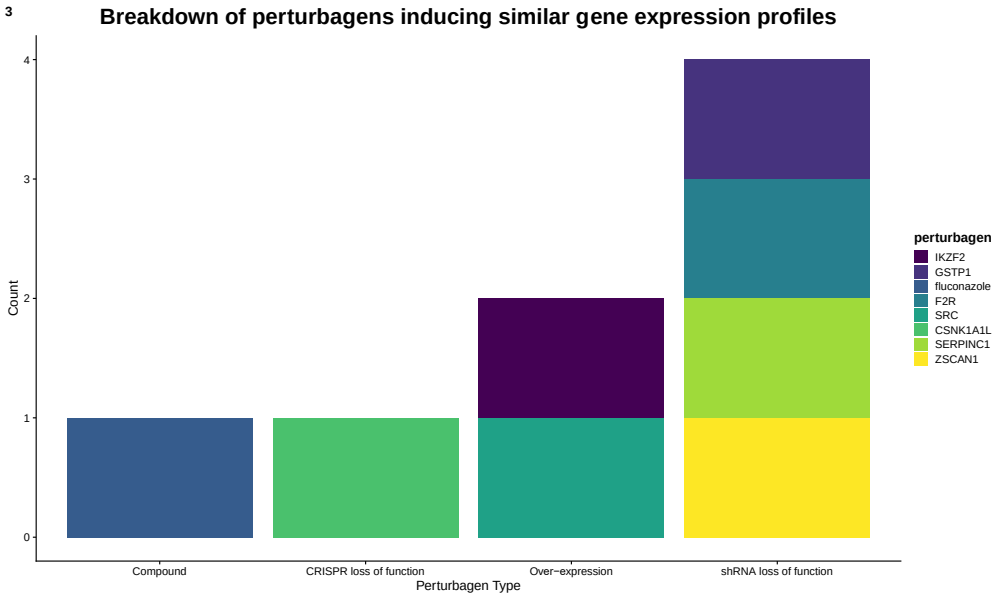


Tableau vs RStudio Drug Repurposing Graphs

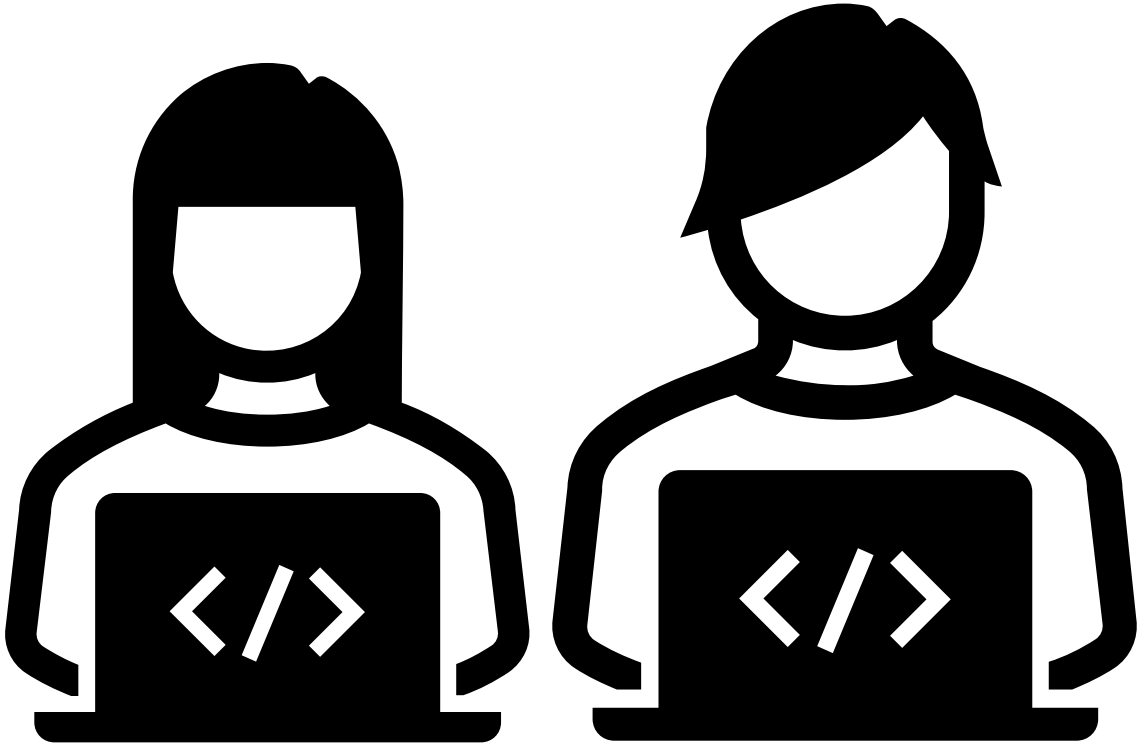






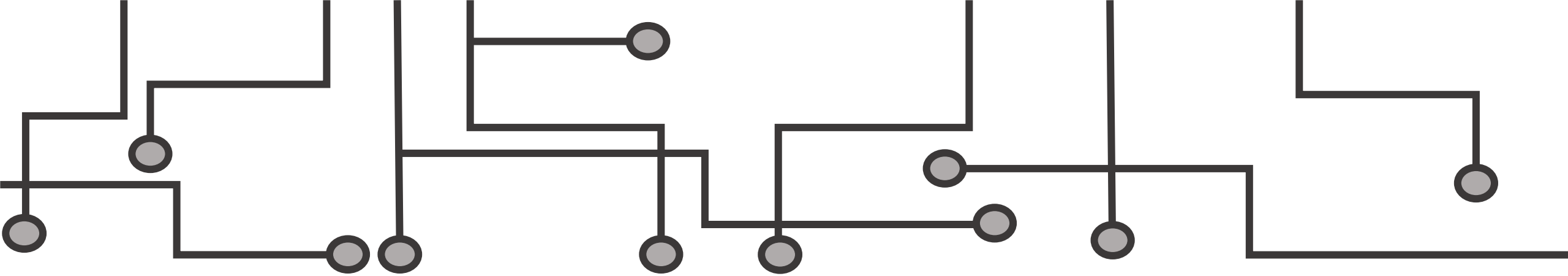
HELPFUL RESOURCES

- ***CLUE/CMap***
- ***ggplot2***
- ***ggplot2/tidyverse***
- ***cow_plot***
- ***Tableau***



Videos

- ***Drug Repurposing Hub at Broad Institute***
- ***CMap Analysis***



THANK YOU FOR ATTENDING!

The Q&A Session will now begin.

Please make sure to fill out the Exit Survey

We value your feedback!

More questions? Please email us at

mmid.coding.workshop@gmail.com or post them to the workshop slack channel