

compare outliers from different comparison cohorts 2023.11.27 11.25.57

hbeale

November 27, 2023

Contents

Define cohort codes	2
Tile plot of all outliers	2
Heatmap shows number of cohorts in which outlier were detected	3
export list of genes found only by TCGA	7
Annotate with combined full low level cohort names	7
How many outliers are present in each combination of cohorts?	8
Tile plot of combination of outliers	8
Annotate with combined full high level cohort names	9
How many outliers are present in each high level combination of cohorts?	10
Annotate with minimal combined cohort abbreviations	10
Annotate with combined cohort abbreviations	11
Summary table for all outliers and low level cohorts	11
Summary table for all outliers and high level cohorts	14
Combined high and low level tables	15
Patient level summary table for all outliers	16
REPEAT ANALYSIS USING ONLY OUTLIERS WITH PATHWAY SUPPORT	16
Tile plot of outliers with pathway support	16
Heatmap shows number of cohorts in which outlier were detected	17
Annotate with combined full cohort names	21
How many outliers with pathway support are present in each combination of cohorts?	21
Patient level summary table for outliers with pathway support	23

Annotate with combined cohort abbreviations 24

Big table of outliers with pathway support 24

version 2023.11.27_11.25.57 adds per-patient analysis

```
underscore_to_space <- function(x) str_replace_all(x, "_", " ")
```

```
outliers <- read_tsv("../input_data/druggable_outliers_from_treehouse_and_other_cohorts_2023_11_09-13_4")
mutate(high_level_cohort = ifelse(str_detect(comparison_cohort, "Treehouse"),
                                   "Treehouse",
                                   comparison_cohort))
```

```
## Rows: 287 Columns: 5
## -- Column specification -----
## Delimiter: "\t"
## chr (4): Sample_ID, comparison_cohort, gene, donor_ID
## lgl (1): pathway_support
##
## i Use `spec()` to retrieve the full column specification for this data.
## i Specify the column types or set `show_col_types = FALSE` to quiet this message.
```

```
n_distinct(outliers$Sample_ID)
```

```
## [1] 34
```

```
n_distinct(outliers$donor_ID)
```

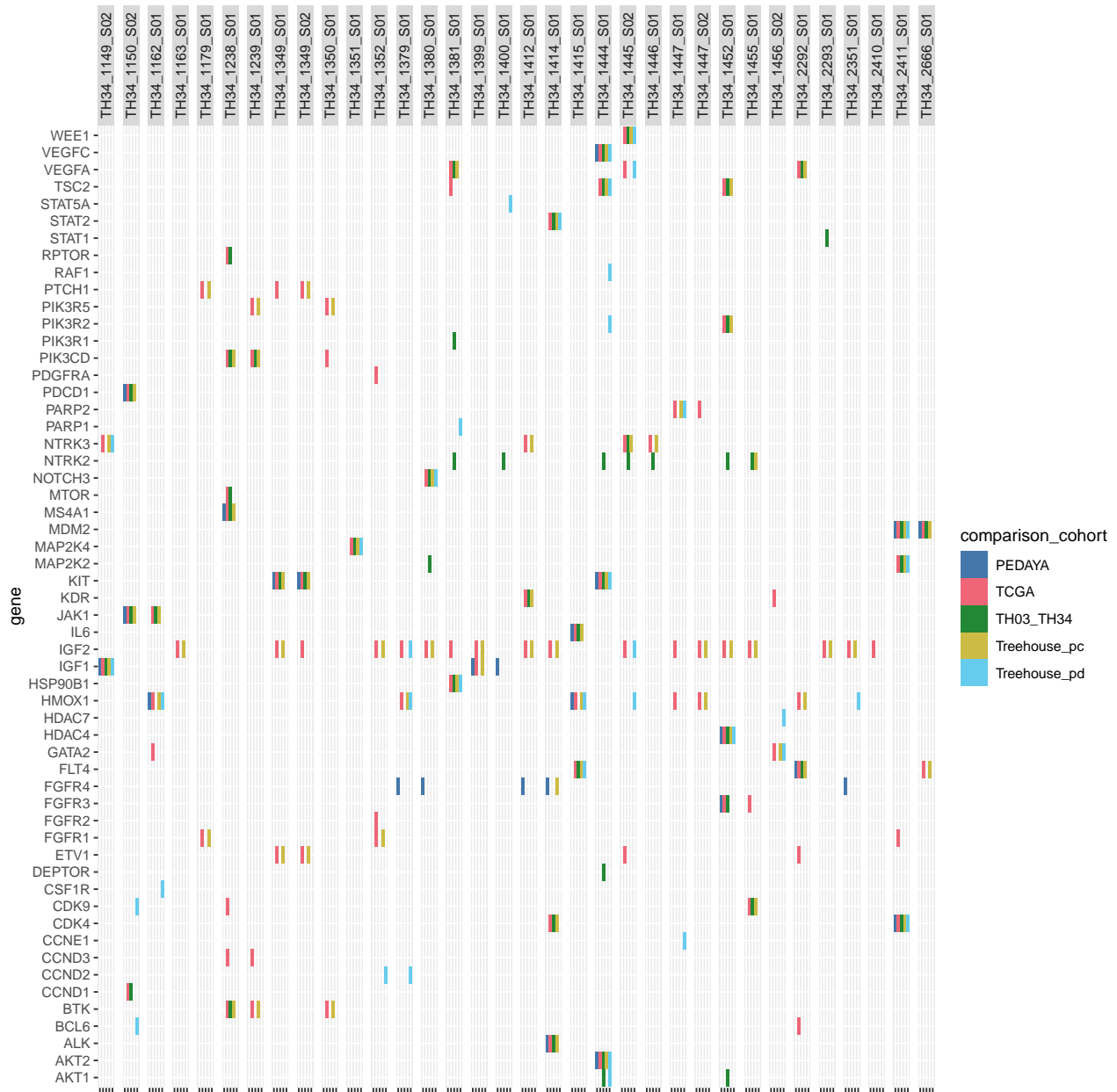
```
## [1] 32
```

Define cohort codes

```
cohort_codes <- tibble(
  cohort_name =
    c("PEDAYA", "TCGA", "TH03_TH34", "Treehouse_pc", "Treehouse_pd"),
  cohort_code =
    c("P", "T", "S", "C", "D"))
```

Tile plot of all outliers

```
ggplot(outliers) +
  geom_tile(aes(x=comparison_cohort,
                y=gene,
                fill = comparison_cohort)) +
  facet_wrap(~Sample_ID,
             nrow = 1) +
  theme(#axis.text.x = element_text(angle = 90, hjust = 1, vjust = 0.5),
        axis.text.x = element_blank(),
        strip.text.x = element_text(angle = 90),
        ) +
  xlab("") +
  scale_fill_bright()
```



Heatmap shows number of cohorts in which outlier were detected

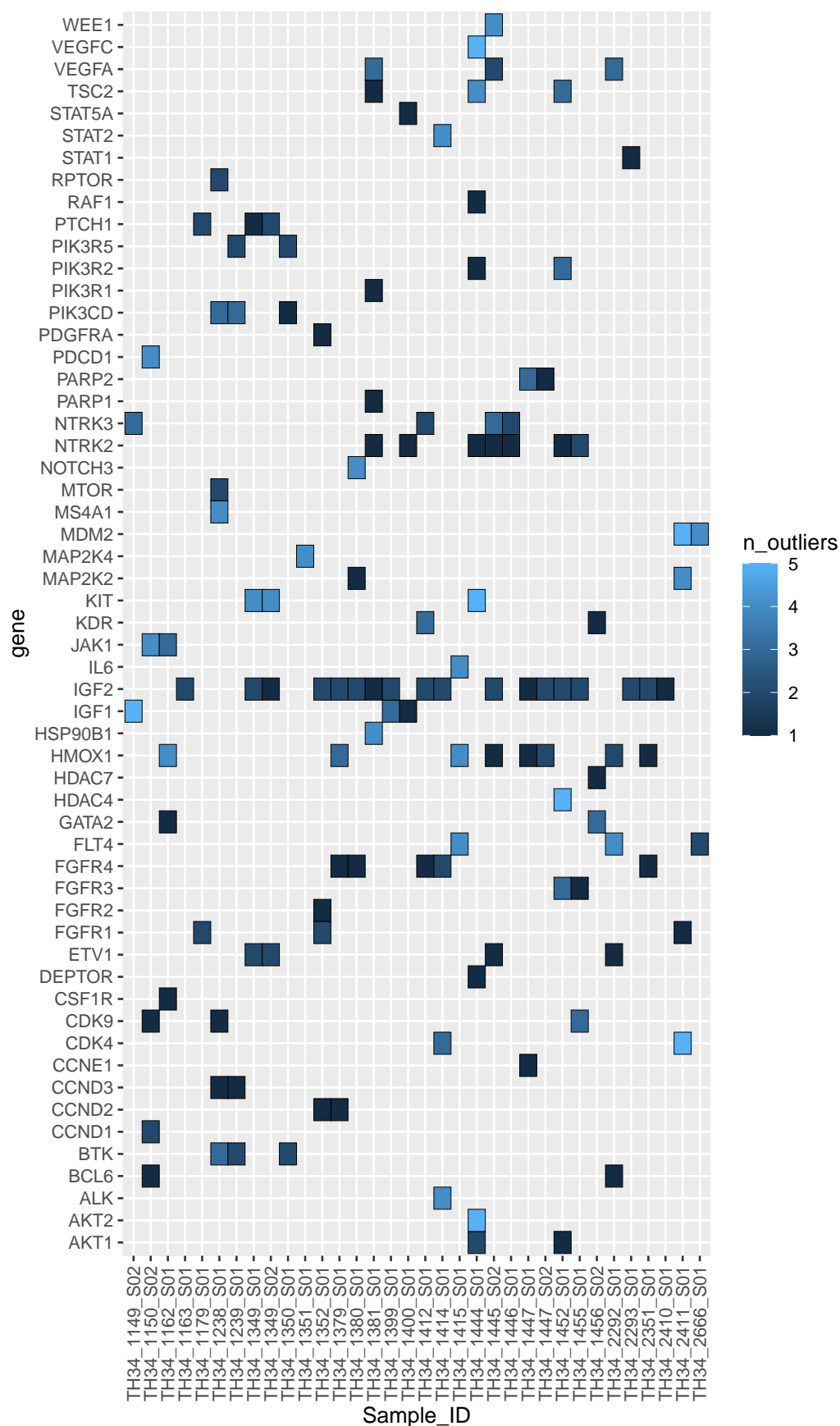
I can make this look better if we decide to use it, but it's non-trivial

```
outliers_heatmap_data <- outliers %>%
  group_by(Sample_ID, gene) %>%
  summarize(n_outliers = n())
```

```
## `summarise()` has grouped output by 'Sample_ID'. You can override using the
## `.groups` argument.
```

```
ggplot(outliers_heatmap_data) +
  geom_tile(aes(x=Sample_ID,
                y=gene,
```

```
        fill = n_outliers),  
        color = "black") +  
#theme_bw() +  
theme(axis.text.x = element_text(angle = 90, hjust = 1, vjust = 0.5))
```

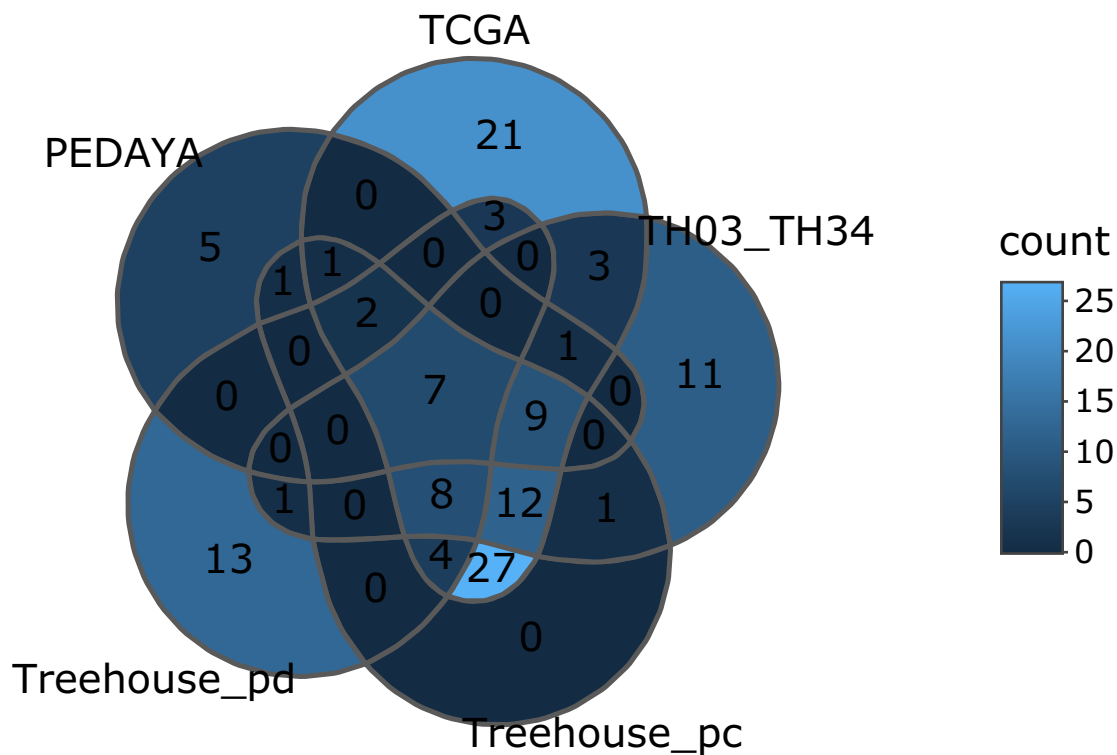


```
library(ggVennDiagram)
raw_outliers_for_venn <- outliers %>%
  mutate(sample_gene = paste(Sample_ID, gene, sep = "_")) %>%
  arrange(comparison_cohort) %>%
  select(sample_gene, comparison_cohort) %>%
  group_split(comparison_cohort)

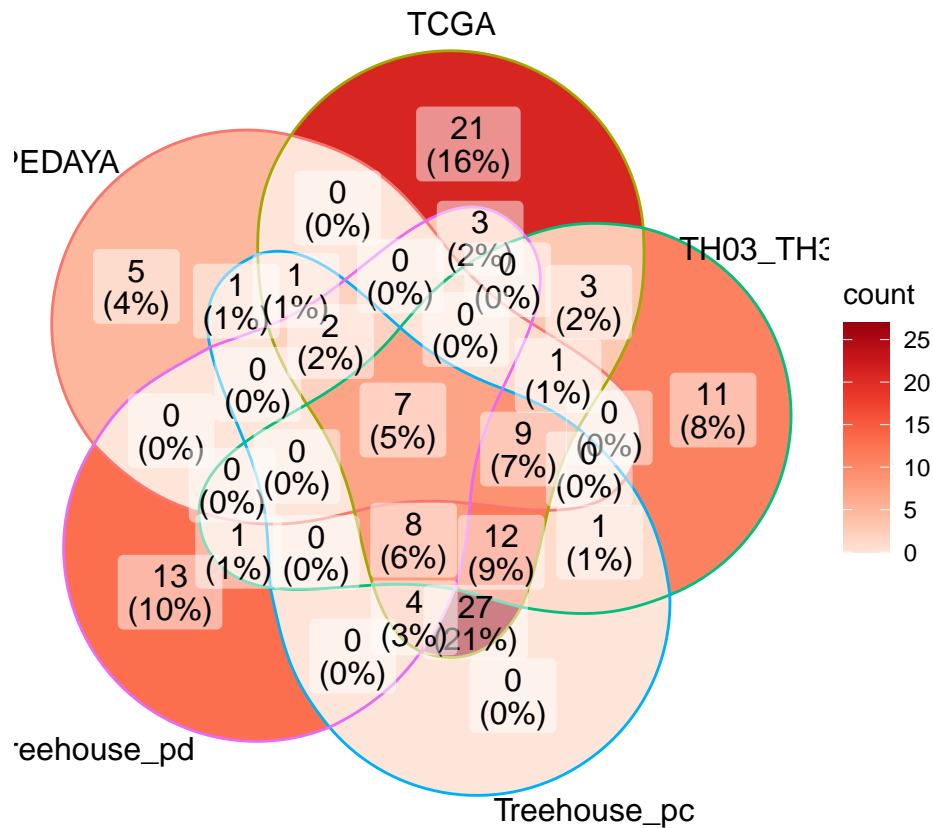
list_of_outliers_for_venn <- lapply(raw_outliers_for_venn, function(x) x %>% pull(sample_gene))
names(list_of_outliers_for_venn) <- unique(outliers$comparison_cohort) %>% sort

ggVennDiagram(list_of_outliers_for_venn,
  show_intersect = TRUE)

## Warning in geom_text(aes_string(label = "count", text = "text"), x =
## label_coord[, : Ignoring unknown aesthetics: text
```



```
ggVennDiagram(list_of_outliers_for_venn) +
  scale_fill_distiller(palette = "Reds", direction = 1)
```



export list of genes found only by TCGA

```
outliers %>%
  group_by(Sample_ID, gene) %>%
  filter(length(comparison_cohort) == 1,
         "TCGA" %in% comparison_cohort) %>%
  ungroup %>%
  select(gene) %>% write_tsv("../gather_input_data/genes found only by TCGA in at least one sample.txt")
```

Annotate with combined full low level cohort names

```
collapse_fun <- function(x){ paste(x, collapse = ", ") }

all_outliers_combined_wide <- outliers %>%
  select(-pathway_support, -donor_ID, -high_level_cohort) %>%
  pivot_wider(names_from = Sample_ID,
              values_from = comparison_cohort,
              values_fn = collapse_fun)

n_distinct(outliers$Sample_ID)
```

```
## [1] 34
```

comparison_cohorts	n	percent
TCGA, Treehouse_pc	27	20.8%
TCGA	21	16.2%
Treehouse_pd	13	10.0%
TCGA, TH03_TH34, Treehouse_pc	12	9.2%
TH03_TH34	11	8.5%
PEDAYA, TCGA, TH03_TH34, Treehouse_pc	9	6.9%
TCGA, TH03_TH34, Treehouse_pc, Treehouse_pd	8	6.2%
PEDAYA, TCGA, TH03_TH34, Treehouse_pc, Treehouse_pd	7	5.4%
PEDAYA	5	3.8%
TCGA, Treehouse_pc, Treehouse_pd	4	3.1%
TCGA, TH03_TH34	3	2.3%
TCGA, Treehouse_pd	3	2.3%
PEDAYA, TCGA, Treehouse_pc, Treehouse_pd	2	1.5%
PEDAYA, TCGA, TH03_TH34	1	0.8%
PEDAYA, TCGA, Treehouse_pc	1	0.8%
PEDAYA, Treehouse_pc	1	0.8%
TH03_TH34, Treehouse_pc	1	0.8%
TH03_TH34, Treehouse_pd	1	0.8%
Total	130	-

```
n_distinct(outliers$gene)
```

```
## [1] 56
```

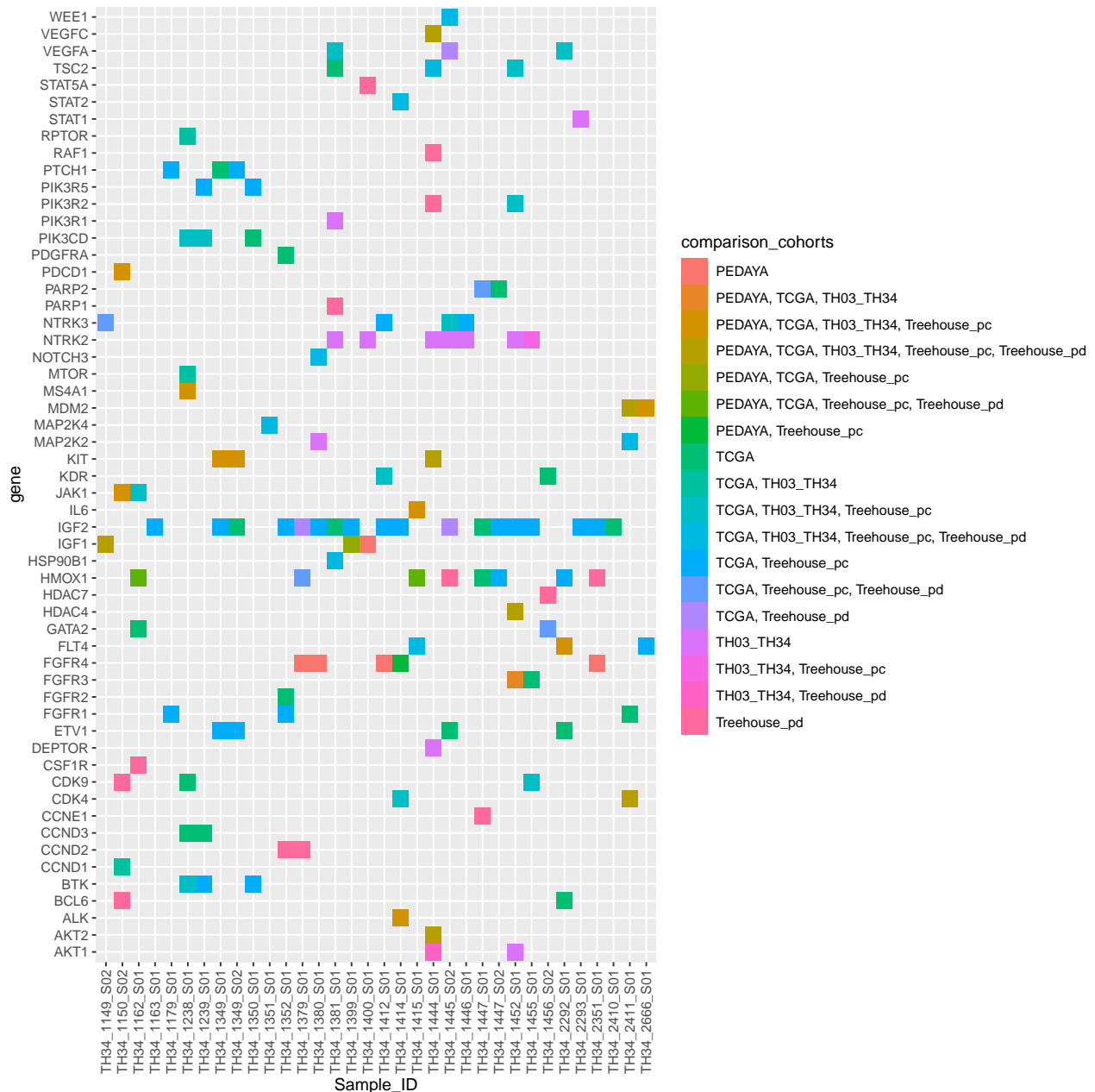
```
all_outliers_combined_long <- all_outliers_combined_wide %>%
  pivot_longer(-gene,
    names_to = "Sample_ID",
    values_to = "comparison_cohorts") %>%
  na.omit()
```

How many outliers are present in each combination of cohorts?

```
tabyl(all_outliers_combined_long,
  comparison_cohorts) %>%
  arrange(desc(n)) %>%
  adorn_pct_formatting() %>%
  adorn_totals() %>%
  kbl() %>%
  kable_styling(full_width = F)
```

Tile plot of combination of outliers

```
ggplot(all_outliers_combined_long) +
  geom_tile(aes(x=Sample_ID,
    y=gene,
    fill = comparison_cohorts)) +
  theme(axis.text.x = element_text(angle = 90, hjust = 1, vjust = 0.5))
```

```
n_distinct(all_outliers_combined_long$Sample_ID)
```

```
## [1] 34
```

Annotate with combined full high level cohort names

```
high_level_all_outliers_combined_wide <- outliers %>%
  select(-pathway_support, -donor_ID, -comparison_cohort) %>%
  distinct() %>%
  pivot_wider(names_from = Sample_ID,
              values_from = high_level_cohort,
              values_fn = collapse_fun)
```

comparison_cohorts	n	percent
TCGA, Treehouse	34	26.2%
TCGA	21	16.2%
TCGA, TH03_TH34, Treehouse	20	15.4%
PEDAYA, TCGA, TH03_TH34, Treehouse	16	12.3%
Treehouse	13	10.0%
TH03_TH34	11	8.5%
PEDAYA	5	3.8%
PEDAYA, TCGA, Treehouse	3	2.3%
TCGA, TH03_TH34	3	2.3%
TH03_TH34, Treehouse	2	1.5%
PEDAYA, TCGA, TH03_TH34	1	0.8%
PEDAYA, Treehouse	1	0.8%
Total	130	-

```
n_distinct(outliers$Sample_ID)

## [1] 34
n_distinct(outliers$gene)

## [1] 56
high_level_all_outliers_combined_long <- high_level_all_outliers_combined_wide %>%
  pivot_longer(-gene,
    names_to = "Sample_ID",
    values_to = "comparison_cohorts") %>%
  na.omit()
```

How many outliers are present in each high level combination of cohorts?

```
tabyl(high_level_all_outliers_combined_long,
  comparison_cohorts) %>%
  arrange(desc(n)) %>%
  adorn_pct_formatting() %>%
  adorn_totals() %>%
  kbl() %>%
  kable_styling(full_width = F)
```

Annotate with minimal combined cohort abbreviations

```
collapse_fun_no_coma <- function(x){ paste(x,collapse = "") }

# backslashes prevent asterisks from being interpreted as italics in the kbl table

all_outliers_min_abbrev_combined_wide <- outliers %>%
  left_join(cohort_codes,
```

```

      by=c("comparison_cohort"="cohort_name")) %>%
mutate(cohort_code_pathway = ifelse(pathway_support,
                                   paste0(cohort_code, "\\*"),
                                   cohort_code)) %>%

select(-pathway_support, -donor_ID,
       -comparison_cohort,
       -cohort_code) %>%
pivot_wider(names_from = Sample_ID,
            values_from = cohort_code_pathway,
            values_fn = collapse_fun_no_coma,
            values_fill = "")

all_outliers_min_abbrev_combined_wide %>%
  arrange(gene) %>%
  rename_all(underscore_to_space) %>%
  kbl() %>%
  kable_styling(full_width = F,
                bootstrap_options = "bordered")

```

Annotate with combined cohort abbreviations

```

all_outliers_abbrev_combined_wide <- outliers %>%
  left_join(cohort_codes,
            by=c("comparison_cohort"="cohort_name")) %>%
  select(-pathway_support, -donor_ID,
         -comparison_cohort) %>%
  pivot_wider(names_from = Sample_ID,
              values_from = cohort_code,
              values_fn = collapse_fun,
              values_fill = "")

all_outliers_abbrev_combined_wide %>%
  arrange(gene) %>%
  rename_all(underscore_to_space) %>%
  kbl() %>%
  kable_styling(full_width = F,
                bootstrap_options = "bordered")

```

Summary table for all outliers and low level cohorts

```

n_outliers_detected_by_any_method <- outliers %>%
  select(Sample_ID, gene) %>%
  distinct %>%
  nrow()

n_outliers_with_pathway_support_detected_by_any_method <- outliers %>%
  filter(pathway_support) %>%
  select(Sample_ID, gene) %>%

```

gene	high level cohort	TH34 1162 S01	TH34 1149 S02	TH34 1238 S01	TH34 1349 S01	TH34 1349 S02
AKT1	TH03_TH34					
AKT1	Treehouse					
AKT2	PEDAYA					
AKT2	TCGA					
AKT2	TH03_TH34					
AKT2	Treehouse					
ALK	PEDAYA					
ALK	TCGA					
ALK	TH03_TH34					
ALK	Treehouse					
BCL6	TCGA					
BCL6	Treehouse					
BTK	TCGA			T*		
BTK	TH03_TH34			S*		
BTK	Treehouse			C*		
CCND1	TCGA					
CCND1	TH03_TH34					
CCND2	Treehouse					
CCND3	TCGA			T*		
CCNE1	Treehouse					
CDK4	PEDAYA					
CDK4	TCGA					
CDK4	TH03_TH34					
CDK4	Treehouse					
CDK9	TCGA			T*		
CDK9	TH03_TH34					
CDK9	Treehouse					
CSF1R	Treehouse	D*				
DEPTOR	TH03_TH34					
ETV1	TCGA				T	T*
ETV1	Treehouse				C*	C*
FGFR1	TCGA					
FGFR1	Treehouse					
FGFR2	TCGA					
FGFR3	PEDAYA					
FGFR3	TCGA					
FGFR3	TH03_TH34					
FGFR4	PEDAYA					
FGFR4	Treehouse					
FLT4	PEDAYA					
FLT4	TCGA					
FLT4	TH03_TH34					
FLT4	Treehouse					
GATA2	TCGA	T				
GATA2	Treehouse					
HDAC4	PEDAYA					
HDAC4	TCGA					
HDAC4	TH03_TH34					
HDAC4	Treehouse					
HDAC7	Treehouse					
HMOX1	PEDAYA	P				
HMOX1	TCGA	T				
HMOX1	Treehouse	CD*				
HSP90B1	TCGA		12			
HSP90B1	TH03_TH34					
HSP90B1	Treehouse					
IGF1	PEDAYA		P*			
IGF1	TCGA		T*			

gene	high level cohort	TH34 1162 S01	TH34 1149 S02	TH34 1238 S01	TH34 1349 S01	TH34 1349 S02
AKT1	TH03_TH34					
AKT1	Treehouse					
AKT2	PEDAYA					
AKT2	TCGA					
AKT2	TH03_TH34					
AKT2	Treehouse					
ALK	PEDAYA					
ALK	TCGA					
ALK	TH03_TH34					
ALK	Treehouse					
BCL6	TCGA					
BCL6	Treehouse					
BTK	TCGA			T		
BTK	TH03_TH34			S		
BTK	Treehouse			C		
CCND1	TCGA					
CCND1	TH03_TH34					
CCND2	Treehouse					
CCND3	TCGA			T		
CCNE1	Treehouse					
CDK4	PEDAYA					
CDK4	TCGA					
CDK4	TH03_TH34					
CDK4	Treehouse					
CDK9	TCGA			T		
CDK9	TH03_TH34					
CDK9	Treehouse					
CSF1R	Treehouse	D				
DEPTOR	TH03_TH34					
ETV1	TCGA				T	T
ETV1	Treehouse				C	C
FGFR1	TCGA					
FGFR1	Treehouse					
FGFR2	TCGA					
FGFR3	PEDAYA					
FGFR3	TCGA					
FGFR3	TH03_TH34					
FGFR4	PEDAYA					
FGFR4	Treehouse					
FLT4	PEDAYA					
FLT4	TCGA					
FLT4	TH03_TH34					
FLT4	Treehouse					
GATA2	TCGA	T				
GATA2	Treehouse					
HDAC4	PEDAYA					
HDAC4	TCGA					
HDAC4	TH03_TH34					
HDAC4	Treehouse					
HDAC7	Treehouse					
HMOX1	PEDAYA	P				
HMOX1	TCGA	T				
HMOX1	Treehouse	C, D				
HSP90B1	TCGA		13			
HSP90B1	TH03_TH34					
HSP90B1	Treehouse					
IGF1	PEDAYA		P			
IGF1	TCGA		T			

comparison cohort	n outliers detected	n outliers with pathway support	pct outliers with pathway support	pct outliers
PEDAYA	26	12	46	
TCGA	98	74	76	
TH03_TH34	53	39	74	
Treehouse_pc	72	47	65	
Treehouse_pd	38	29	76	
Total	130	101	78	

```

distinct %>%
  nrow()
# these have pathway support in at least one cohort

outlier_summary <- outliers %>%
  group_by(comparison_cohort) %>%
  summarize(n_outliers_detected = n(),
            n_outliers_with_pathway_support = sum(pathway_support),
            pct_outliers_with_pathway_support = 100*n_outliers_with_pathway_support/n_outliers_detected,
            pct_outliers_detected = 100*n_outliers_detected/n_outliers_detected_by_any_method)

outlier_summary_with_totals <-
  bind_rows(outlier_summary,
            tibble(comparison_cohort= " Total",
                  n_outliers_detected = n_outliers_detected_by_any_method,
                  n_outliers_with_pathway_support = n_outliers_with_pathway_support_detected_by_any_method,
                  pct_outliers_with_pathway_support = 100*n_outliers_with_pathway_support_detected_by_any_method))

outlier_summary_with_totals %>%
  rename_all(underscore_to_space) %>%
  kbl(digits = c(NA, 0, 0, 0, 0)) %>%
  kable_styling(full_width = F)

```

Summary table for all outliers and high level cohorts

```

n_outliers_detected_by_any_method <- outliers %>%
  select(Sample_ID, gene) %>%
  distinct %>%
  nrow()

n_outliers_with_pathway_support_detected_by_any_method <- outliers %>%
  filter(pathway_support) %>%
  select(Sample_ID, gene) %>%
  distinct %>%
  nrow()
# these have pathway support in at least one cohort

high_level_outlier_summary <- outliers %>%

```

high level cohort	n outliers detected	n outliers with pathway support	pct outliers with pathway support	pct outliers
Treehouse	89	64	72	
TH03_TH34	53	39	74	
TCGA	98	74	76	
PEDAYA	26	12	46	
Total	130	101	78	

```

group_by(high_level_cohort, Sample_ID, gene) %>%
  summarize(pathway_support = any(pathway_support)) %>%
  group_by(high_level_cohort) %>%
  summarize(n_outliers_detected = n(),
            n_outliers_with_pathway_support = sum(pathway_support),
            pct_outliers_with_pathway_support = 100*n_outliers_with_pathway_support/n_outliers_detected,
            pct_outliers_detected = 100*n_outliers_detected/n_outliers_detected_by_any_method)

## `summarise()` has grouped output by 'high_level_cohort', 'Sample_ID'. You can
## override using the `.groups` argument.

high_level_outlier_summary_with_totals <-
bind_rows(high_level_outlier_summary %>%
  arrange(desc(high_level_cohort)),
  tibble(high_level_cohort= " Total",
        n_outliers_detected = n_outliers_detected_by_any_method,
        n_outliers_with_pathway_support = n_outliers_with_pathway_support_detected_by_any_method,
        pct_outliers_with_pathway_support = 100*n_outliers_with_pathway_support_detected_by_any_method)

high_level_outlier_summary_with_totals %>%
  rename_all(underscore_to_space) %>%
  kbl(format.args = list(big.mark = ","), digits = c(NA, 0, 0, 0, 0)) %>%
  kable_styling(full_width = F)

```

Combined high and low level tables

```

high_low <- bind_rows(
  high_level_outlier_summary_with_totals %>%
    rename(comparison_cohort=high_level_cohort) %>%
    mutate(index = c(1, 4:7)),
  outlier_summary_with_totals %>%
    filter(str_detect(comparison_cohort, "Treehouse")) %>%
    mutate(index = 2:3)
) %>%
  arrange(index) %>%
  select(-index)

high_low %>%
  rename_all(underscore_to_space) %>%
  kbl(format.args = list(big.mark = ","), digits = c(NA, 0, 0, 0, 0)) %>%
  kable_styling(full_width = F)

```

comparison cohort	n outliers detected	n outliers with pathway support	pct outliers with pathway support	pct outliers
Treehouse	89	64	72	
Treehouse_pc	72	47	65	
Treehouse_pd	38	29	76	
TH03_TH34	53	39	74	
TCGA	98	74	76	
PEDAYA	26	12	46	
Total	130	101	78	

name	n_patients_with_druggable_outliers
PEDAYA	18
TCGA	31
TH03_TH34	22
Treehouse	31
Treehouse_pc	30
Treehouse_pd	18

Patient level summary table for all outliers

```

outliers %>%
  group_by(donor_ID) %>%
  summarize(any_PEDAYA = "PEDAYA" %in% comparison_cohort,
            any_TH03_TH34 = "TH03_TH34" %in% comparison_cohort,
            any_TCGA = "TCGA" %in% comparison_cohort,
            any_Treehouse_pc = "Treehouse_pc" %in% comparison_cohort,
            any_Treehouse_pd = "Treehouse_pd" %in% comparison_cohort,
            any_Treehouse = any_Treehouse_pc | any_Treehouse_pd) %>%
  pivot_longer(starts_with("any")) %>%
  mutate(name = str_remove(name, "any_")) %>%
  group_by(name) %>%
  summarize(n_patients_with_druggable_outliers = sum(value)) %>%
  kbl() %>%
  kable_styling(full_width = F)

```

REPEAT ANALYSIS USING ONLY OUTLIERS WITH PATHWAY SUPPORT

Tile plot of outliers with pathway support

```

ggplot(outliers %>%
  filter(pathway_support)) +
  geom_tile(aes(x=comparison_cohort,
               y=gene,
               fill = comparison_cohort)) +
  facet_wrap(~Sample_ID,

```



```

nrow = 1) +
theme(#axis.text.x = element_text(angle = 90, hjust = 1, vjust = 0.5),
      axis.text.x = element_blank(),
      strip.text.x = element_text(angle = 90),
    ) +
xlab("") +
scale_fill_bright()

```



Heatmap shows number of cohorts in which outlier were detected

I can make this look better if we decide to use it, but it's non-trivial

```

pathway_outliers_heatmap_data <- outliers %>%
  filter(pathway_support) %>%
  group_by(Sample_ID, gene) %>%
  summarize(n_outliers = n())

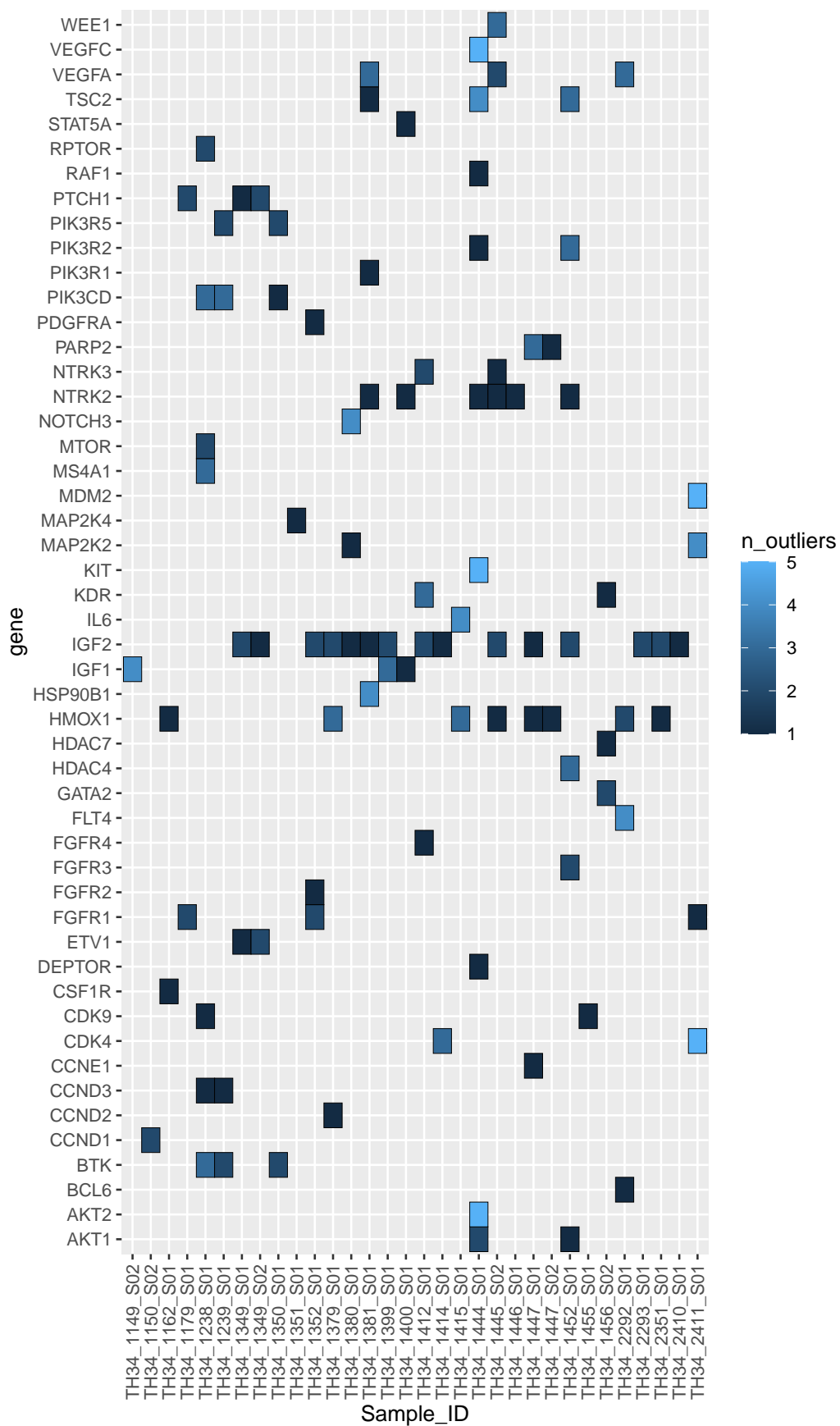
```

`summarise()` has grouped output by 'Sample_ID'. You can override using the
` .groups ` argument.

```

ggplot(pathway_outliers_heatmap_data) +
  geom_tile(aes(x=Sample_ID,
                y=gene,
                fill = n_outliers),
            color = "black") +
  #theme_bw() +
  theme(axis.text.x = element_text(angle = 90, hjust = 1, vjust = 0.5))

```



```

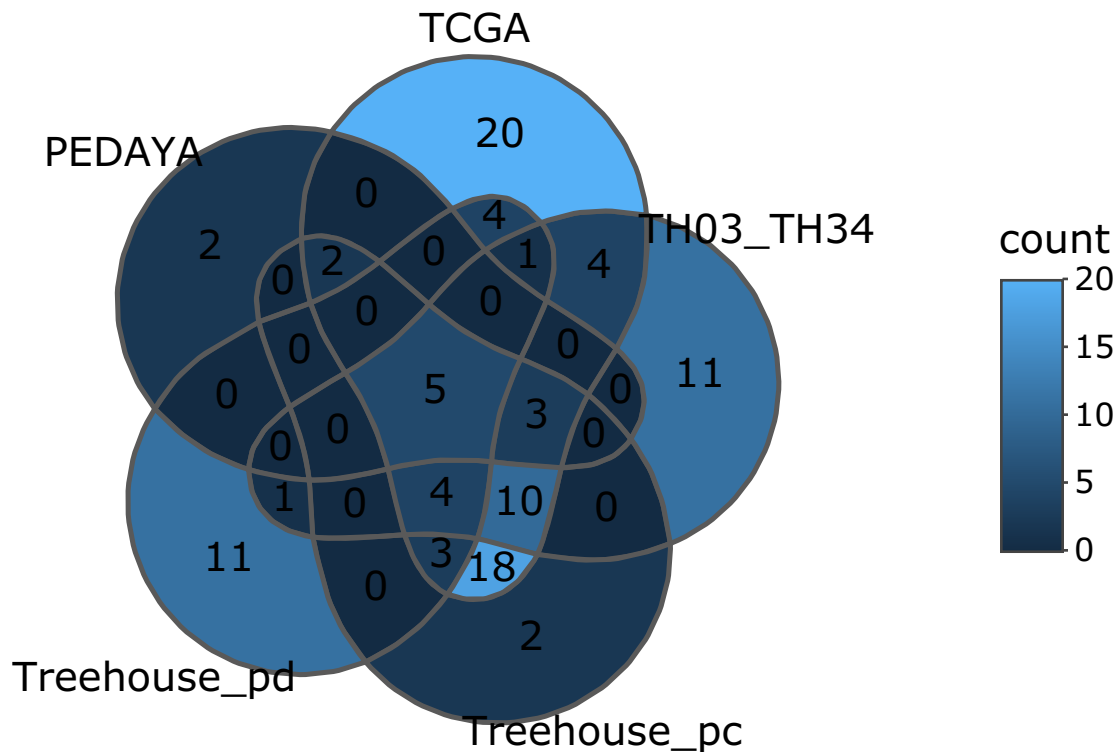
raw_pathway_support_outliers_for_venn <- outliers %>%
  filter(pathway_support) %>%
  mutate(sample_gene = paste(Sample_ID, gene, sep = "_")) %>%
  arrange(comparison_cohort) %>%
  select(sample_gene, comparison_cohort) %>%
  group_split(comparison_cohort)

list_of_pathway_support_outliers_for_venn <- lapply(raw_pathway_support_outliers_for_venn, function(x)
names(list_of_pathway_support_outliers_for_venn) <- outliers %>%
  filter(pathway_support) %>%
  arrange(comparison_cohort) %>%
  select(comparison_cohort) %>%
  distinct() %>%
  pull(comparison_cohort)

ggVennDiagram(list_of_pathway_support_outliers_for_venn,
              show_intersect = TRUE)

## Warning in geom_text(aes_string(label = "count", text = "text"), x =
## label_coord[, : Ignoring unknown aesthetics: text

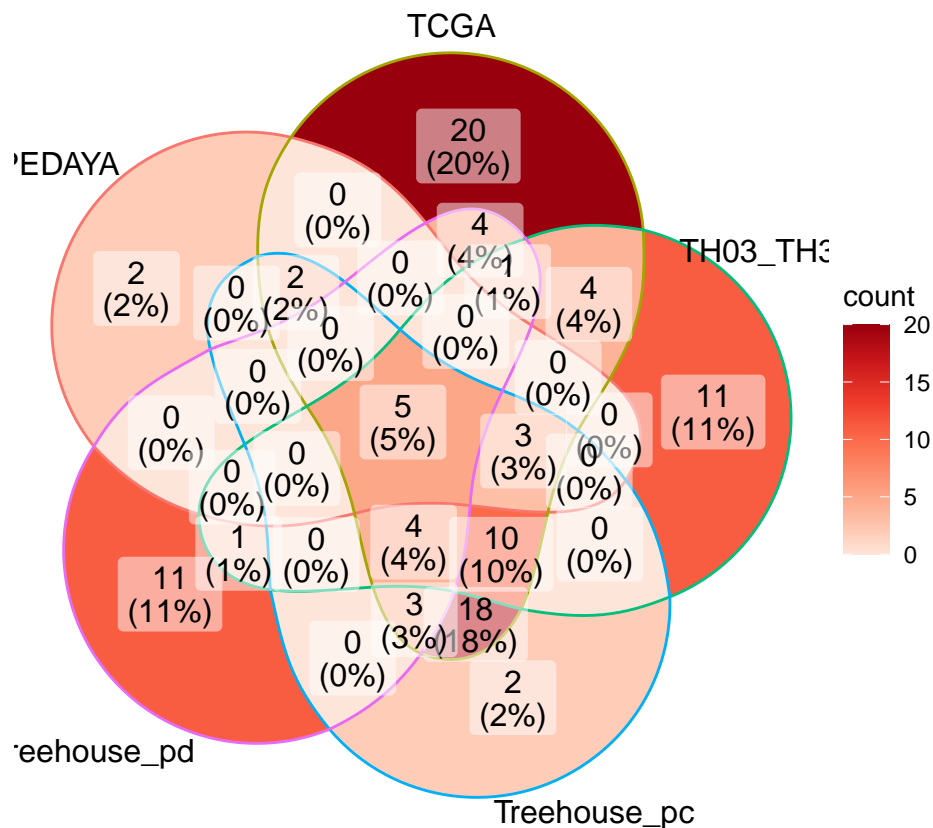
```



```

ggVennDiagram(list_of_pathway_support_outliers_for_venn) +
  scale_fill_distiller(palette = "Reds", direction = 1)

```



Annotate with combined full cohort names

```
outliers_with_pathway_support_combined_wide <- outliers %>%
  filter(pathway_support) %>%
  select(-pathway_support, -donor_ID) %>%
  pivot_wider(names_from = Sample_ID,
              values_from = comparison_cohort,
              values_fn = collapse_fun)

outliers_with_pathway_support_combined_long <- outliers_with_pathway_support_combined_wide %>%
  pivot_longer(-gene,
              names_to = "Sample_ID",
              values_to = "comparison_cohorts") %>%
  na.omit()
```

How many outliers with pathway support are present in each combination of cohorts?

```
tabyl(outliers_with_pathway_support_combined_long,
      comparison_cohorts) %>%
  arrange(desc(n)) %>%
  adorn_pct_formatting() %>%
```

comparison_cohorts	n	percent
TCGA	112	37.1%
TH03_TH34	67	22.2%
Treehouse	37	12.3%
Treehouse_pc	35	11.6%
PEDAYA	22	7.3%
Treehouse_pd	17	5.6%
Treehouse_pc, Treehouse_pd	12	4.0%
Total	302	-

```

adorn_totals() %>%
kbl() %>%
kable_styling(full_width = F)

ggplot(outliers_with_pathway_support_combined_long) +
  geom_tile(aes(x=Sample_ID,
                y=gene,
                fill = comparison_cohorts)) +
  theme(axis.text.x = element_text(angle = 90, hjust = 1, vjust = 0.5))

```



```
n_distinct(outliers_with_pathway_support_combined_long$Sample_ID)
```

```
## [1] 33
```

Patient level summary table for outliers with pathway support

```
outliers %>%
  filter(pathway_support) %>%
  group_by(donor_ID) %>%
  summarize(any_PEDAYA = "PEDAYA" %in% comparison_cohort,
            any_TH03_TH34 = "TH03_TH34" %in% comparison_cohort,
            any_TCGA = "TCGA" %in% comparison_cohort,
```

name	n_patients_with_druggable_outliers
PEDAYA	8
TCGA	26
TH03_TH34	16
Treehouse	26
Treehouse_pc	22
Treehouse_pd	13

```

any_Treehouse_pc = "Treehouse_pc" %in% comparison_cohort,
any_Treehouse_pd = "Treehouse_pd" %in% comparison_cohort,
any_Treehouse = any_Treehouse_pc | any_Treehouse_pd) %>%
pivot_longer(starts_with("any")) %>%
mutate(name = str_remove(name, "any_")) %>%
group_by(name) %>%
summarize(n_patients_with_druggable_outliers = sum(value)) %>%
kbl() %>%
kable_styling(full_width = F)

```

Annotate with combined cohort abbreviations

```

outliers_with_pathway_support_abbrev_combined_wide <- outliers %>%
  filter(pathway_support) %>%
  left_join(cohort_codes,
            by=c("comparison_cohort"="cohort_name")) %>%
  select(-pathway_support, -donor_ID,
         -comparison_cohort) %>%
  pivot_wider(names_from = Sample_ID,
              values_from = cohort_code,
              values_fn = collapse_fun,
              values_fill = "")

```

Big table of outliers with pathway support

```

outliers_with_pathway_support_abbrev_combined_wide %>%
  arrange(gene) %>%
  rename_all(underscore_to_space) %>%
  kbl() %>%
  kable_styling(full_width = F,
                bootstrap_options = "bordered")

```

```
sessionInfo()
```

```

## R version 4.2.1 (2022-06-23)
## Platform: aarch64-apple-darwin20 (64-bit)
## Running under: macOS Monterey 12.2
##
## Matrix products: default
## BLAS: /Library/Frameworks/R.framework/Versions/4.2-arm64/Resources/lib/libRblas.0.dylib
## LAPACK: /Library/Frameworks/R.framework/Versions/4.2-arm64/Resources/lib/libRlapack.dylib

```


gene	high level cohort	TH34 1149 S02	TH34 1238 S01	TH34 1399 S01	TH34 1400 S01	TH34 1412 S01
AKT1	TH03_TH34					
AKT1	Treehouse					
AKT2	PEDAYA					
AKT2	TCGA					
AKT2	TH03_TH34					
AKT2	Treehouse					
BCL6	TCGA					
BTK	TCGA		T			
BTK	TH03_TH34		S			
BTK	Treehouse		C			
CCND1	TCGA					
CCND1	TH03_TH34					
CCND2	Treehouse					
CCND3	TCGA		T			
CCNE1	Treehouse					
CDK4	PEDAYA					
CDK4	TCGA					
CDK4	TH03_TH34					
CDK4	Treehouse					
CDK9	TCGA		T			
CSF1R	Treehouse					
DEPTOR	TH03_TH34					
ETV1	TCGA					
ETV1	Treehouse					
FGFR1	TCGA					
FGFR1	Treehouse					
FGFR2	TCGA					
FGFR3	TCGA					
FGFR3	TH03_TH34					
FGFR4	PEDAYA					P
FLT4	PEDAYA					
FLT4	TCGA					
FLT4	TH03_TH34					
FLT4	Treehouse					
GATA2	TCGA					
GATA2	Treehouse					
HDAC4	TCGA					
HDAC4	TH03_TH34					
HDAC4	Treehouse					
HDAC7	Treehouse					
HMOX1	PEDAYA					
HMOX1	TCGA					
HMOX1	Treehouse					
HSP90B1	TCGA					
HSP90B1	TH03_TH34					
HSP90B1	Treehouse					
IGF1	PEDAYA	P		P	P	
IGF1	TCGA	T		T		
IGF1	TH03_TH34	S				
IGF1	Treehouse	C		C		
IGF2	TCGA			T		T
IGF2	Treehouse			C		C
IL6	PEDAYA					
IL6	TCGA		25			
IL6	TH03_TH34					
IL6	Treehouse					
KDR	TCGA					T
KDR	TH03_TH34					C

```

##
## locale:
## [1] en_US.UTF-8/en_US.UTF-8/en_US.UTF-8/C/en_US.UTF-8/en_US.UTF-8
##
## attached base packages:
## [1] stats      graphics  grDevices  utils      datasets  methods   base
##
## other attached packages:
## [1] ggVennDiagram_1.2.2 cowplot_1.1.1      gridExtra_2.3
## [4] kableExtra_1.3.4    khroma_1.10.0      janitor_2.1.0
## [7] forcats_0.5.2       stringr_1.5.0      dplyr_1.0.10
## [10] purrr_0.3.5         readr_2.1.3        tidyr_1.2.1
## [13] tibble_3.2.1        ggplot2_3.4.4      tidyverse_1.3.2
##
## loaded via a namespace (and not attached):
## [1] fs_1.6.3            sf_1.0-9            lubridate_1.9.0
## [4] bit64_4.0.5         RColorBrewer_1.1-3 webshot_0.5.4
## [7] httr_1.4.4          tools_4.2.1         backports_1.4.1
## [10] utf8_1.2.3          R6_2.5.1            KernSmooth_2.23-20
## [13] DBI_1.1.3           lazyeval_0.2.2      colorspace_2.1-0
## [16] withr_2.5.0         tidyselect_1.2.0    processx_3.8.0
## [19] bit_4.0.5           compiler_4.2.1      cli_3.6.1
## [22] rvest_1.0.3         xml2_1.3.3          plotly_4.10.1
## [25] labeling_0.4.2      scales_1.2.1        classInt_0.4-9
## [28] callr_3.7.3         proxy_0.4-27        systemfonts_1.0.4
## [31] digest_0.6.33       yulab.utils_0.0.6   rmarkdown_2.23
## [34] svglite_2.1.0       pkgconfig_2.0.3     htmltools_0.5.5
## [37] dbplyr_2.2.1        fastmap_1.1.1       highr_0.10
## [40] htmlwidgets_1.6.2   rlang_1.1.1         readxl_1.4.1
## [43] rstudioapi_0.14     farver_2.1.1        generics_0.1.3
## [46] jsonlite_1.8.7      crosstalk_1.2.0     vroom_1.6.0
## [49] googlesheets4_1.0.1 magrittr_2.0.3      Rcpp_1.0.11
## [52] munsell_0.5.0       fansi_1.0.4         lifecycle_1.0.3
## [55] stringi_1.7.12      yaml_2.3.7          snakecase_0.11.0
## [58] grid_4.2.1          parallel_4.2.1      crayon_1.5.2
## [61] haven_2.5.1         hms_1.1.2           ps_1.7.2
## [64] knitr_1.43          pillar_1.9.0        reprex_2.0.2
## [67] glue_1.6.2          evaluate_0.21       data.table_1.14.6
## [70] modelr_0.1.10       vctrs_0.6.3         tzdb_0.3.0
## [73] cellranger_1.1.0    gtable_0.3.3        assertthat_0.2.1
## [76] xfun_0.39           broom_1.0.1         e1071_1.7-13
## [79] class_7.3-20        googledrive_2.0.0   RVenn_1.1.0
## [82] viridisLite_0.4.2   gargle_1.2.1        units_0.8-1
## [85] timechange_0.1.1    ellipsis_0.3.2

```