Main Figures

Selcen Ari Alper Yilmaz

2021-02-18

Figures

This file consists analysis steps and codes of figures in manuscript.

```
mirtarbase_node_perturbations <- readRDS("mirtarbase_node_perturbations.RDS")
# data includes results of find_node_perturbation()
# function for tumor and normal samples of each
# patient.
```

Venn diagrams Figure 4A

```
gene_dist <- mirtarbase_node_perturbations %>% group_by(name) %>%
    summarise(perturbed = sum(perturbed_count > 0,
        na.rm = TRUE)) %>% ungroup()



effective_nodes <- mirtarbase_node_perturbations %>%
    left_join(gene_dist, by = "name") %>% filter(perturbed >
    10, perturbed_count > 78) %>% distinct(name)
```

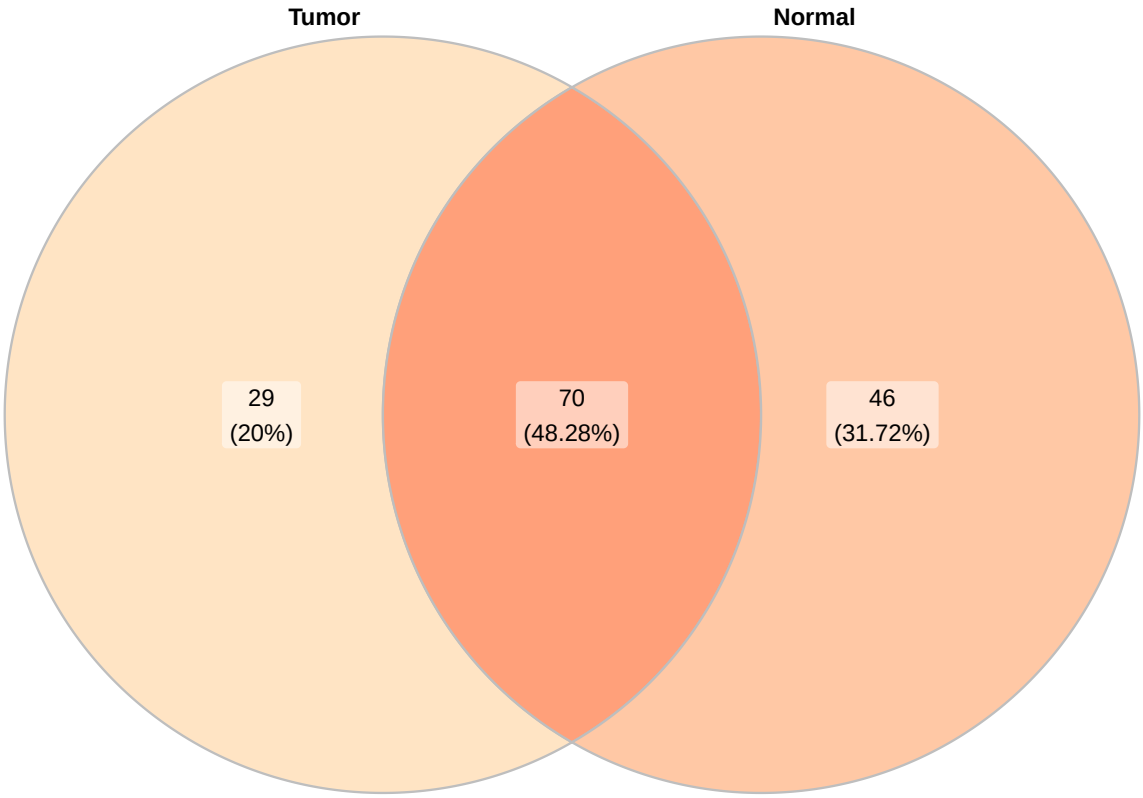
- Detecting highly perturbing gene nodes in tumor tissues

```
lim_1_gene_tumor <- mirtarbase_node_perturbations %>%
    left_join(gene_dist, by = "name") %>% mutate(tissue_type = ifelse(endsWith(file_name,
    "01A"), "Tumor", "Normal"), node_type = ifelse(name %in%
    effective_nodes$name, "effective", "non_effective")) %>%
    filter(node_type == "effective", perturbed_count >
        78, perturbed > 10) %>% filter(tissue_type !=
    "Normal", startsWith(name, "ENS")) %>% distinct(name) %>%
    pull()
```

- Detecting highly perturbing gene nodes in normal tissues

```
lim_1_gene_normal <- mirtarbase_node_perturbations %>%
    left_join(gene_dist, by = "name") %>% mutate(tissue_type = ifelse(endsWith(file_name,
    "01A"), "Tumor", "Normal"), node_type = ifelse(name %in%
    effective_nodes$name, "effective", "non_effective")) %>%
    filter(node_type == "effective", perturbed_count >
        78, perturbed > 10) %>% filter(tissue_type ==
    "Normal", startsWith(name, "ENS")) %>% distinct(name) %>%
    pull()
```

Venn diagram of genes

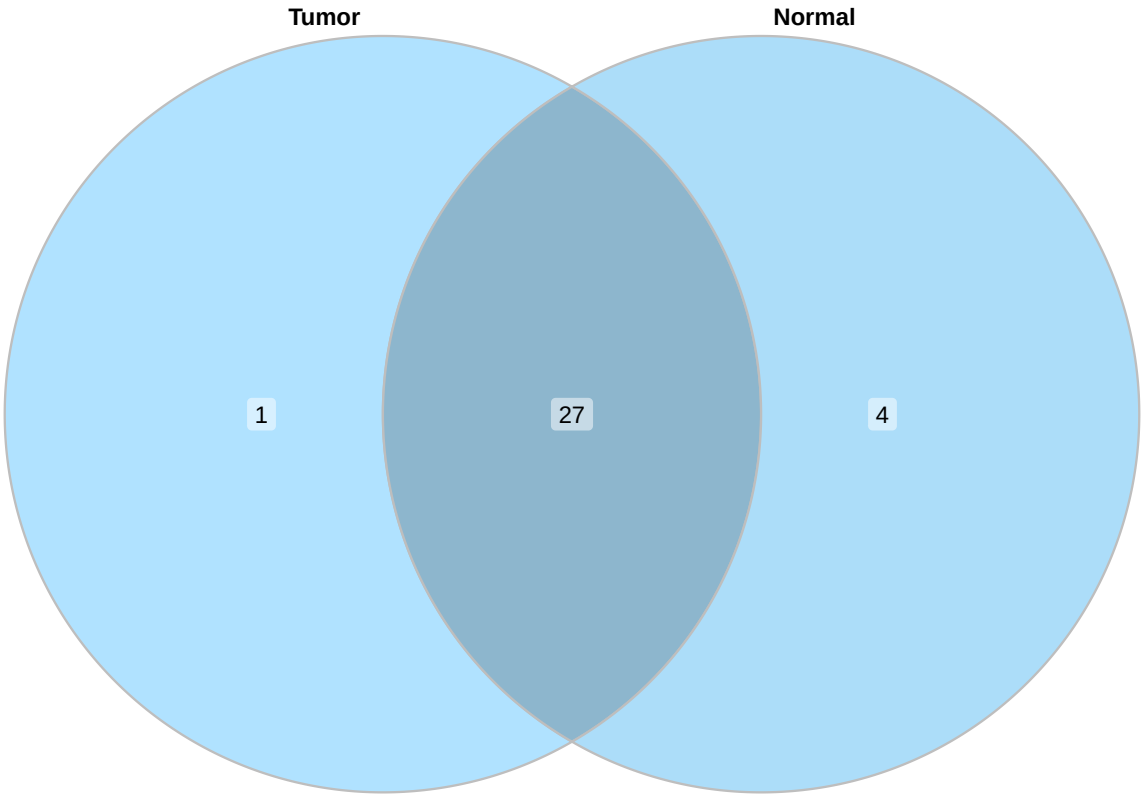


- Detecting highly perturbing mirna nodes in tumor tissues

```
lim_1_mirna_tumor <- mirtarbase_node_perturbations %>%
    left_join(gene_dist, by = "name") %>% mutate(tissue_type = ifelse(endsWith(file_name,
    "01A"), "Tumor", "Normal"), node_type = ifelse(name %in%
    effective_nodes$name, "effective", "non_effective")) %>%
    filter(node_type == "effective", perturbed_count >
        78, perturbed > 10) %>% filter(tissue_type !=
    "Normal", startsWith(name, "hsa")) %>% distinct(name) %>%
    pull()
```

- Detecting highly perturbing mirna nodes in normal tissues

```
lim_1_mirna_normal <- mirtarbase_node_perturbations %>%
    left_join(gene_dist, by = "name") %>% mutate(tissue_type = ifelse(endsWith(file_name,
    "01A"), "Tumor", "Normal"), node_type = ifelse(name %in%
    effective_nodes$name, "effective", "non_effective")) %>%
    filter(node_type == "effective", perturbed_count >
        78, perturbed > 10) %>% filter(tissue_type ==
    "Normal", startsWith(name, "hsa")) %>% distinct(name) %>%
    pull()
```




```
library(ggpubr)

# pA <- grid.arrange(pa1, pa2)

pA <- ggpubr::ggarrange(pa1, pa2, nrow = 2)
pA
# ggsave(filename = 'lim1_PE_gene_mirtarbase.svg',
# width = 6, height = 6)
```

Functional Annotation Analysis

```
Anno_tumor_specific <- readRDS("Anno_tumor_specific.RDS")  # includes functional annotation of 29 genes
```

#Fuctional annotation of Tumor specific 29 genes. **Figure 4B**

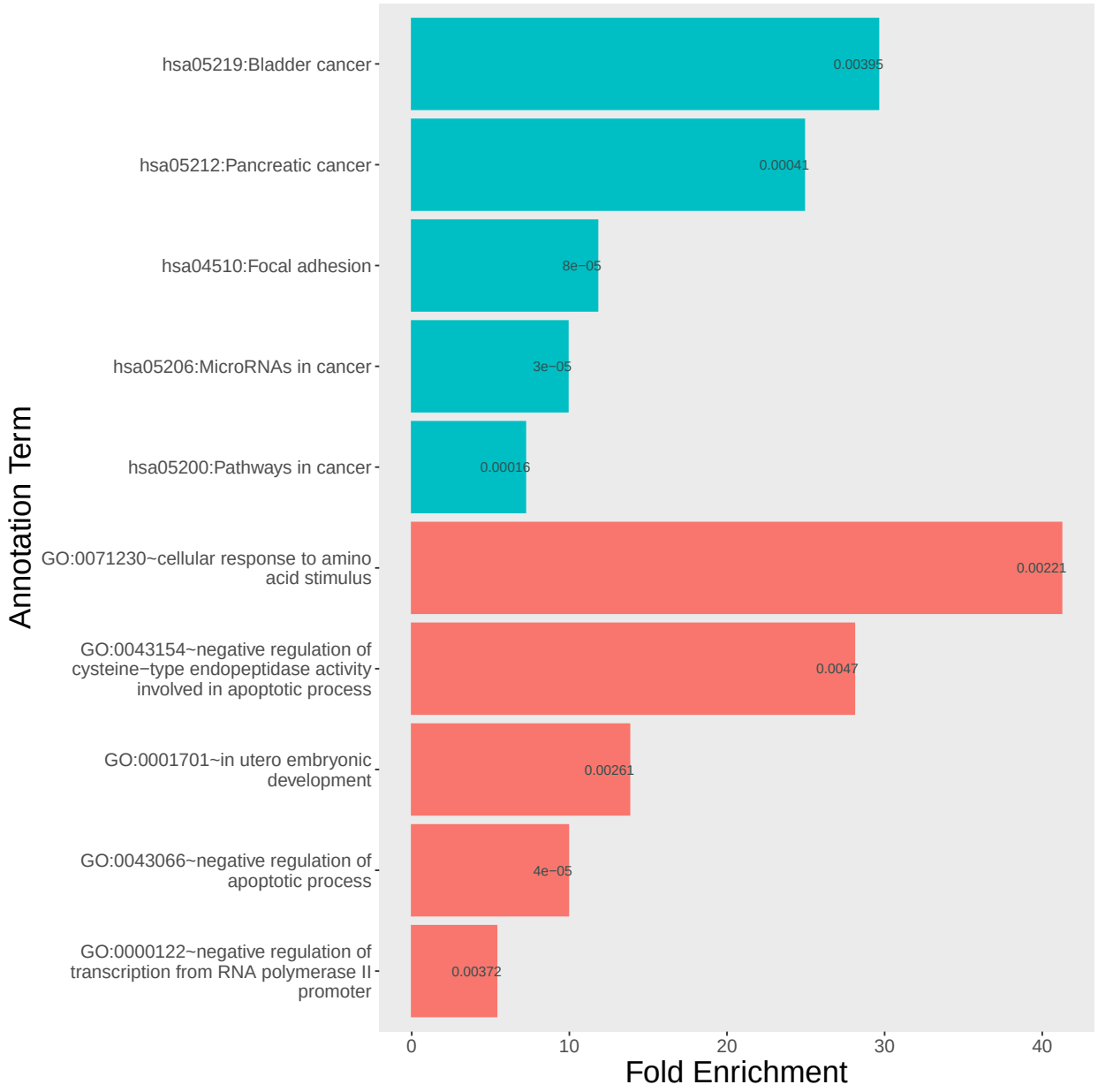
Tumor only:

```
tumor_only_annotation <- Anno_tumor_specific %>% filter(startsWith(Category,
    c("KEGG", "GO")), !str_detect(Category, "CC_DIRECT")) %>%
    filter(!str_detect(Category, "MF_DIRECT")) %>%
    clean_names() %>% group_by(category) %>% filter(row_number() <
    6)


position <- rev(c("hsa05219:Bladder cancer", "hsa05212:Pancreatic cancer",
    "hsa04510:Focal adhesion", "hsa05206:MicroRNAs in cancer",
    "hsa05200:Pathways in cancer", "GO:0071230~cellular response to amino acid stimulus",
    "GO:0043154~negative regulation of cysteine-type endopeptidase activity involved in apoptotic proces
    "GO:0001701~in utero embryonic development", "GO:0043066~negative regulation of apoptotic process",
    "GO:0000122~negative regulation of transcription from RNA polymerase II promoter"))



tumor_only_annotation %>% group_by(category) %>% arrange(-desc(fold_enrichment))
```

```
## # A tibble: 10 x 13
## # Groups:   category [2]
##    category term  count percent p_value genes list_total pop_hits pop_total
##    <chr>    <chr> <dbl>   <dbl>   <dbl> <chr>      <dbl>    <dbl>     <dbl>
##  1 GOTERM_~ GO:0~     6    20.7 3.72e-3 ENSG~         26      720     16792
##  2 KEGG_PA~ hsa0~     7    24.1 1.64e-4 ENSG~         17      393      6879
##  3 KEGG_PA~ hsa0~     7    24.1 2.76e-5 ENSG~         17      286      6879
##  4 GOTERM_~ GO:0~     7    24.1 4.38e-5 ENSG~         26      455     16792
##  5 KEGG_PA~ hsa0~     6    20.7 7.65e-5 ENSG~         17      206      6879
##  6 GOTERM_~ GO:0~     4    13.8 2.61e-3 ENSG~         26      187     16792
##  7 KEGG_PA~ hsa0~     4    13.8 4.13e-4 ENSG~         17       65      6879
##  8 GOTERM_~ GO:0~     3    10.3 4.70e-3 ENSG~         26       69     16792
##  9 KEGG_PA~ hsa0~     3    10.3 3.95e-3 ENSG~         17       41      6879
## 10 GOTERM_~ GO:0~     3    10.3 2.21e-3 ENSG~         26       47     16792
## # ... with 4 more variables: fold_enrichment <dbl>, bonferroni <dbl>,
## #   benjamini <dbl>, fdr <dbl>
```

```
tumor_only_annotation %>% ggplot(aes(x = fold_enrichment,
    y = term)) + geom_col(aes(color = category, fill = category)) +
    geom_text(aes(label = round(p_value, 5)), hjust = 0.9,
        color = "darkslategrey", size = 3) + xlab("Fold Enrichment") +
    ylab("Annotation Term") + theme(panel.grid = element_blank(),
    axis.title = element_text(size = 20), axis.text = element_text(size = 12),
    legend.position = "none") + scale_y_discrete(labels = scales::wrap_format(40),
```

```
    position = "left", limits = position)

ggsave(filename = "mirtarbase_annotation.svg", width = 8,
    height = 6)
```



This figure was removed from manuscript. ## Network construction with overall functional annotation: Additional figure important genes from functional Annotation

```
significant_node_graph <- readRDS("significant_node_graph.RDS")
# consists network of highly perturbing nodes.
hg19 <- readRDS("hg19.RDS")  #For id matching (obtained via bioMaRt package)
```

Annotation network:

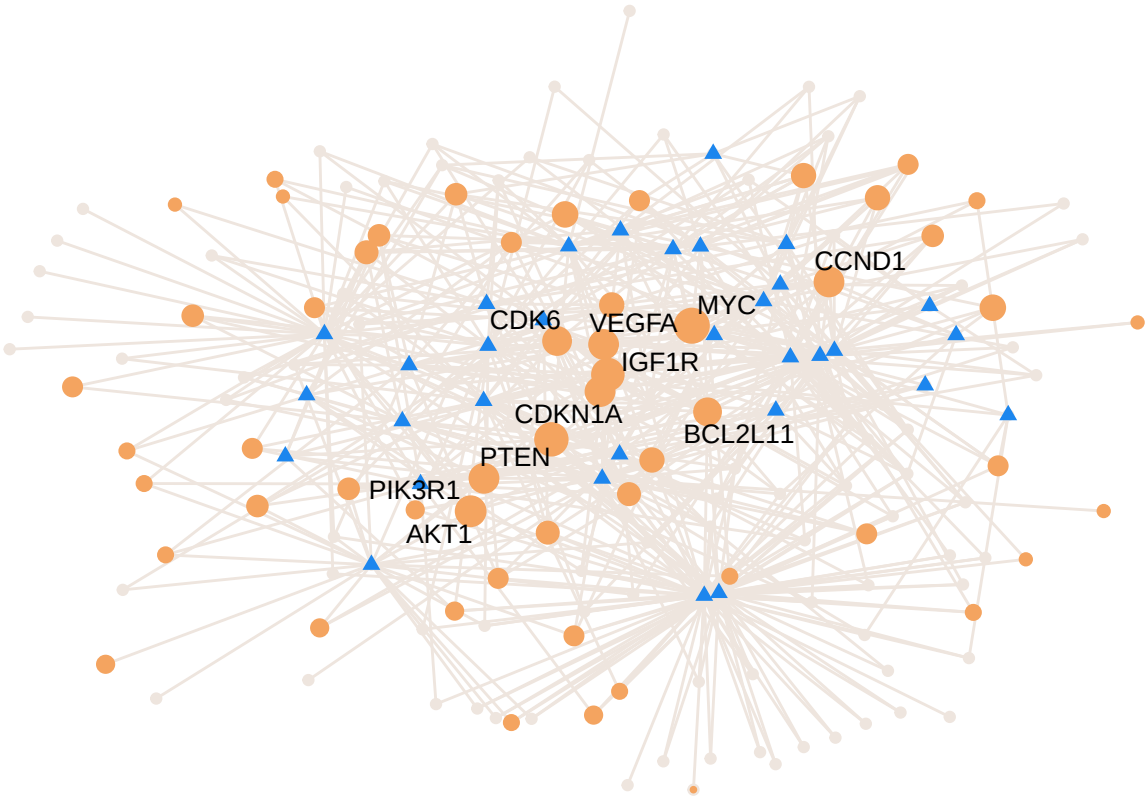
```
Annotation_overall <- readRDS("Annotation_overall.RDS")
# consists functional annotation of 145 high
# perturbing nodes.

gene_top_annotation_graph <- Annotation_overall %>%
    filter(startsWith(Category, c("KEGG", "GO")), !str_detect(Category,
        "CC_DIRECT")) %>% filter(!str_detect(Category,
    "MF_DIRECT")) %>% clean_names() %>% group_by(category) %>%
    filter(row_number() < 6) %>% dplyr::select(term,
    Gene_Count = count, p_value, genes) %>% mutate(genes2 = str_split(genes,
    ", ")) %>% unnest() %>% ungroup %>% mutate(genes2 = str_trim(genes2,
    side = "both"), genes2 = ifelse(endsWith(genes2,
    ","), substr(genes2, 0, (nchar(genes2) - 1)), genes2),
    interaction_type = term, Ensembl_Gene_Id = genes2,
    competing_name = genes2) %>% dplyr::select(interaction = term,
    Ensembl_Gene_Id, interaction_type, competing_name)
```

```
## Adding missing grouping variables: `category`
```

```
## Warning: `cols` is now required when using unnest().
## Please use `cols = c(genes2)`
```

```
top_important <- significant_node_graph %>% bind_rows(gene_top_annotation_graph) %>%
    as_tbl_graph() %>% mutate(type = ifelse(startsWith(name,
    "hsa"), "KEGG", "GO"), type = ifelse(startsWith(name,
    "hsa-"), "miRNA", type), type = ifelse(startsWith(name,
    "ENSG"), "Gene", type)) %>% mutate(centrality = centrality_degree(mode = "all"),
    annotation = ifelse(name %in% gene_top_annotation_graph$Ensembl_Gene_Id,
        "top_annotation", "other")) %>% as_tibble() %>%
    filter(annotation == "top_annotation", centrality >
        10) %>% dplyr::select(name) %>% pull()


annotation_graph <- significant_node_graph %>% bind_rows(gene_top_annotation_graph) %>%
    as_tbl_graph() %>% mutate(type = ifelse(startsWith(name,
    "hsa"), "KEGG", "GO"), type = ifelse(startsWith(name,
    "hsa-"), "miRNA", type), type = ifelse(startsWith(name,
    "ENSG"), "Gene", type)) %>% mutate(centrality = centrality_degree(mode = "all"),
    annotation = ifelse(name %in% gene_top_annotation_graph$Ensembl_Gene_Id,
        "Enriched Genes", "other")) %>% left_join(hg19,
    by = c(name = "ensembl_gene_id")) %>% filter(type %in%
    c("miRNA", "Gene"))

annotation_graph %>% ggraph(layout = "kk") + geom_edge_link(colour = "seashell2") +
    geom_node_point(color = "seashell2") + geom_node_point(aes(filter = annotation ==
    "Enriched Genes", size = centrality, color = "Enriched Genes"),
    shape = 16) + geom_node_point(aes(filter = type ==
    "miRNA", color = "miRNA"), shape = 17, size = 2) +
    geom_node_point(aes(filter = type == "GO", size = centrality,
        color = "GO"), shape = 18) + geom_node_point(aes(filter = type ==
```

```
    "KEGG", size = centrality, color = "KEGG"), shape = 15) +
    theme_graph(base_family = "sans") + geom_node_text(aes(filter = name %in%
    top_important, label = hgnc_symbol), size = 3.5,
    repel = TRUE) + theme(plot.margin = margin(0, 0,
    0, 0, "cm"), legend.position = "none") + guides(size = FALSE,
    shape = FALSE) + scale_color_manual(name = "Node types",
    values = c(`Enriched Genes` = "sandybrown", miRNA = "dodgerblue2",
        GO = "green", KEGG = "red"))
```



Note: Alternatively, network can also be visualized by using Cytoscape (v 3.8.2) and Rcy3 package from Bioconductor.

```
annotation_graph
# BiocManager::install('RCy3')

library(RCy3)

# After starting cytoscape at desktop, following
# commands are run:

RCy3::cytoscapePing()
createNetworkFromIgraph(annotation_graph)
```

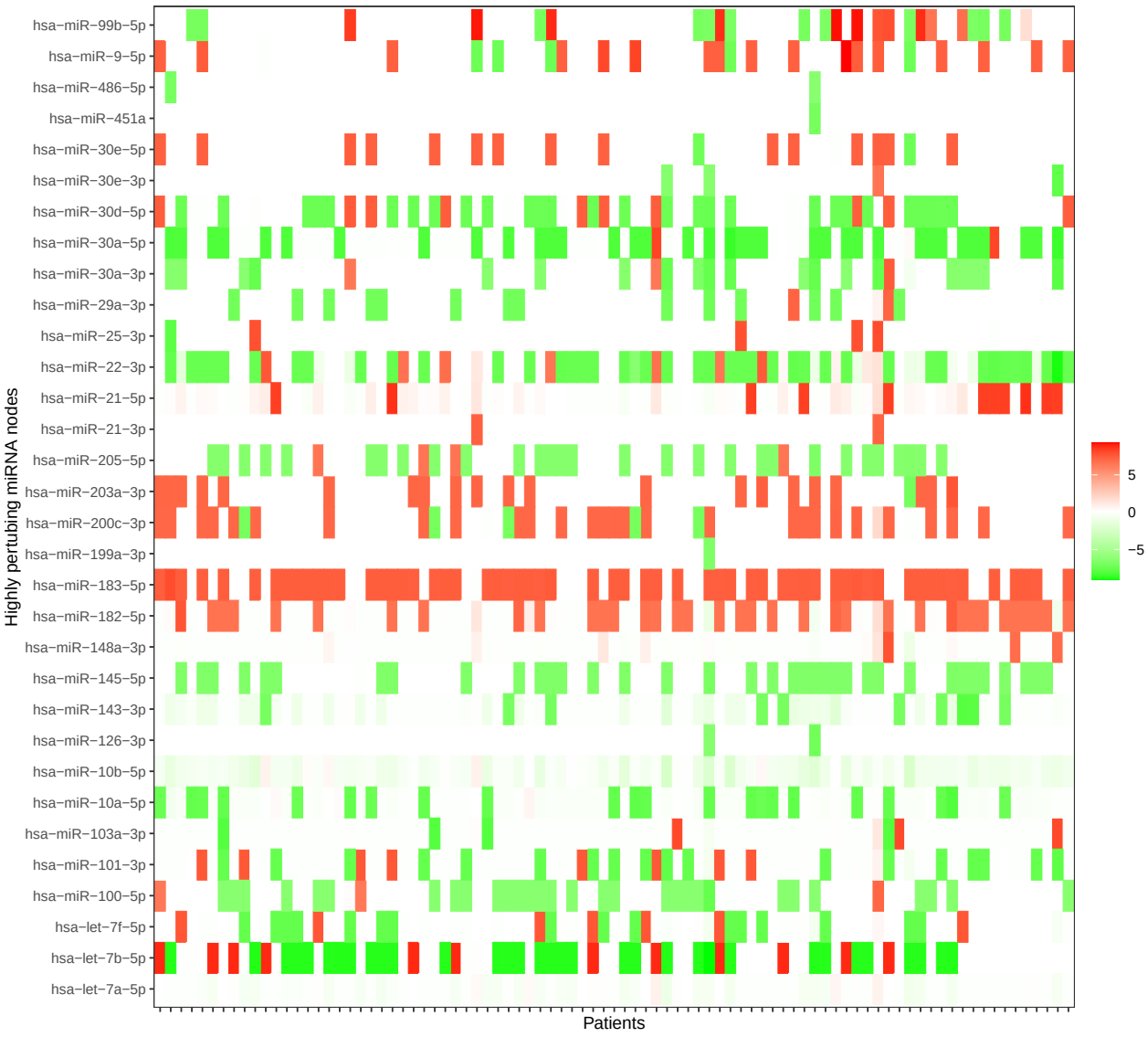
Heat-maps for miRNAs

+Detecting perturbing miRNAs

```
mirna_tumor <- mirtarbase_node_perturbations %>% left_join(gene_dist,
    by = "name") %>% mutate(tissue_type = ifelse(endsWith(file_name,
    "01A"), "Tumor", "Normal"), node_type = ifelse(name %in%
    effective_nodes$name, "effective", "non_effective")) %>%
    filter(node_type == "effective", perturbed_count >
        78, perturbed > 10) %>% filter(tissue_type !=
    "Normal", startsWith(name, "hsa")) %>% mutate(file_name = str_remove(file_name,
    "-01A")) %>% dplyr::select(file_name, name, tissue_type,
    perturbed_count)

mirna_normal <- mirtarbase_node_perturbations %>% left_join(gene_dist,
    by = "name") %>% mutate(tissue_type = ifelse(endsWith(file_name,
    "01A"), "Tumor", "Normal"), node_type = ifelse(name %in%
    effective_nodes$name, "effective", "non_effective")) %>%
    filter(node_type == "effective", perturbed_count >
        78, perturbed > 10) %>% filter(tissue_type ==
    "Normal", startsWith(name, "hsa")) %>% mutate(file_name = str_remove(file_name,
    "-11A")) %>% dplyr::select(file_name, name, tissue_type,
    perturbed_count)
```

Heat map Figure 4C

```
mirna_tumor%>%
  full_join(mirna_normal, by =c("name", "file_name"), suffix= c("_tumor", "_normal"), fill.na = 0)%>%
  dplyr::select(-3, -5)%>%
  mutate(perturbed_count_tumor = ifelse(is.na(perturbed_count_tumor), 1, perturbed_count_tumor),
         perturbed_count_normal = ifelse(is.na(perturbed_count_normal), 1, perturbed_count_normal),
         log_FC= log2(perturbed_count_tumor/perturbed_count_normal))%>%
  ggplot(aes(x = file_name, y = name, fill = log_FC))+
  geom_tile(aes(colour = log_FC, fill = log_FC))+
  theme_test()+
  theme(axis.text.x = element_blank(), #element_text(angle = 90, vjust = 0, hjust=0)
        plot.title = element_text(hjust = 0.5))+
  scale_colour_gradientn(colours = c("green","white", "red"),
                         aesthetics = c("colour", "fill"), na.value = "grey")+
  ylab("Highly pertubing miRNA nodes")+
  xlab("Patients")+
  theme( legend.title = element_blank(), panel.background = element_rect(fill = "white"),
         plot.margin = margin(0, 0, 0, 0, "cm"))

ggsave(filename = 'mirna_comparison.svg', width = 8, height = 4)
```



Heat map Figure 4D

```
mirna_tumor%>%
  full_join(mirna_normal, by =c("name", "file_name"), suffix= c("_tumor", "_normal"), fill.na = 0)%>%
  dplyr::select(-3, -5)%>%
  pivot_longer(cols = c("perturbed_count_tumor", "perturbed_count_normal"),
               names_to = "situ", values_to = "perturbed_count")%>%
  mutate(perturbed_count_log = log2(perturbed_count),
         situ = str_replace(situ, "perturbed_count_normal", 'Perturbed count in normal'),
         situ = str_replace(situ, "perturbed_count_tumor", 'Perturbed count in tumor'),
         perturbed_count = as.numeric(str_replace_na(perturbed_count, 0)),
         perturbed_count_log = as.numeric(str_replace_na(perturbed_count_log, 0)))%>%
  ggplot(aes(x = file_name, y = name, fill = perturbed_count_log))+
  geom_tile(aes(colour = perturbed_count_log, fill = perturbed_count_log))+
  theme_test()+
  theme(axis.text.x = element_blank(), #element_text(angle = 90, vjust = 0, hjust=0)
```

```
        plot.title = element_text(hjust = 0.5))+
  scale_colour_gradientn(colours = c("green","red", "black"), aesthetics = c("colour", "fill"), na.valu
  ylab("Highly pertubing miRNA nodes")+
  xlab("Patients")+
  facet_grid(rows = "situ")+
  theme( legend.title = element_blank(), panel.background = element_rect(fill = "gray93"),
         plot.margin = margin(0, 0, 0, 0, "cm"))


ggsave(filename = 'mirna_comparison_normal_tumor.svg', width = 8, height = 7)
```