

# FIG\_5

Compiled: May 15, 2022

```
library(grid)
library(RColorBrewer)
library(ggplot2)
library(pheatmap)
library(reshape2)
library(ggpubr)
library(dplyr)
library(ggVennDiagram)
library(VennDiagram)
```

## Fig. 5a - Venn diagram: event condition

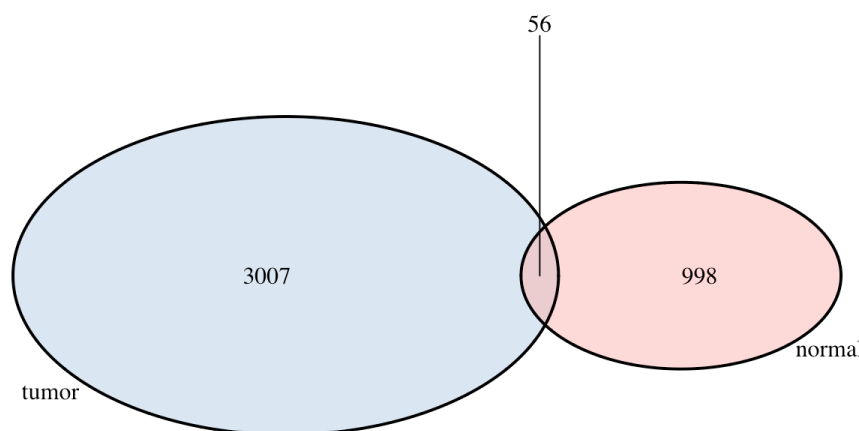
```
df3c <- read.csv("v2.variants.ALL_INDV.annot_hyergeom.linkage_prevalence.txt", sep = "\t", header = F)
# TH179 chr2:232573301:C>T germline 1.0 nan 0.0 0 3
names(df3c) <- c("IND", "var", "gene", "label", "p_val", "tumor_prev", "normal_prev", "tumor_n",
  "normal_n", "group", "group_merged")
head(df3c)
```

IND	var	gene	label	p_val	tumor_prev	normal_prev	tumor_n	normal_n	group
TH238	1034967-1034968	chr1:[1034865,1034890)/.	ENSG00000131591	C1orf159	1	0	1	0	TUMC
TH238	1309771-1309772	chr1:[1309971,1310083)/.	ENSG00000175756	AURKAIP1	1	0	1	0	TUMC
TH238	1309705-1309706	chr1:[1310171,1310376)/.	ENSG00000175756	AURKAIP1	1	0	1	0	TUMC
TH238	1337365-1337366	chr1:[1337982,1341182)/.	ENSG00000242485	MRPL20	1	0	1	0	TUMC
TH238	1565888-1565889	chr1:[1564691,1564763)/.	ENSG00000197530	MIB2	1	0	1	0	TUMC
TH238	1565888-1565889	chr1:[1564946,1565017)/.	ENSG00000197530	MIB2	1	0	1	0	TUMC

```
PAL = brewer.pal(2, "Pastel1")

df4c <- table(df3c$group)

venn.plot <- draw.pairwise.venn(area1 = df4c[2] + df4c[1], area2 = df4c[3] + df4c[1], cross.area = df4c[1],
  category <- c("normal", "tumor"), fill = c(PAL[1], PAL[2]))
```

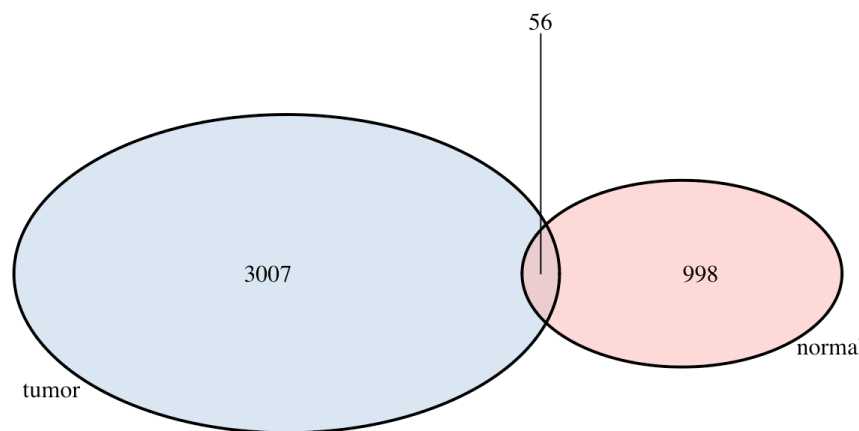


```
mdf3c <- melt(df3c, id.vars = c("IND", "var", "gene", "label", "p_val", "tumor_n", "normal_n",
  "group", "group_merged"))
g1 <- ggplot(mdf3c, aes(x = value)) + geom_histogram(binwidth = 0.9) + facet_wrap(variable ~ group,
  ncol = 3, scales = "free") + theme_classic()

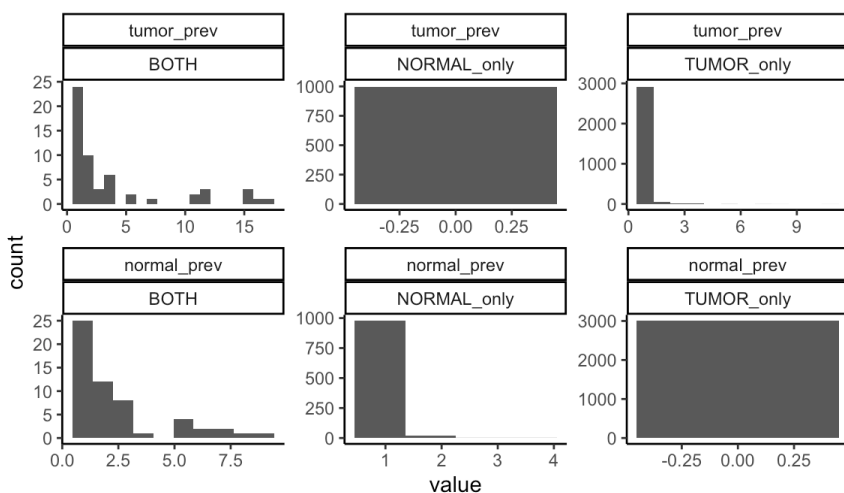
pdf("/Users/giovanni/hoffman_folder/micro_indel_project/FIGS/5a.pdf")
grid.newpage()
grid.draw(venn.plot)
print(g1)
dev.off()
```

```
## quartz_off_screen
##                2
```

```
grid.newpage()
grid.draw(venn.plot)
```



```
print(g1)
```



### Fig. 5b - Count of differential/condition-

specific events

```
df3c <- read.csv("v2.variants.ALL_INDV.annot_hyergeom.linkage_differential.txt", sep = "\t", header = F)
names(df3c) <- c("IND", "var", "gene", "label", "p_val", "event", "condition", "fraction_normal",
  "fraction_tumor", "pval", "group", "n_sig", "n_test")
head(df3c)
```

IND	var	gene	label	p_val	event	condition	fraction_normal	fraction_tumor	p
TH238	1034967-1034968	ENSG00000131591	C1orf159	NaN	chr1:[1034865,1034890)/.	NaN	0	0.5	

IND	var	gene	label	p_val	event	condition	fraction_normal	fraction_tumor	p
TH238	1309771-1309772	ENSG00000175756	AURKAIP1	NaN	chr1:[1309971,1310083)/.	NaN	0	0.5	
TH238	1309705-1309706	ENSG00000175756	AURKAIP1	NaN	chr1:[1310171,1310376)/.	NaN	0	0.5	
TH238	1337365-1337366	ENSG00000242485	MRPL20	NaN	chr1:[1337982,1341182)/.	NaN	0	0.5	
TH238	1565888-1565889	ENSG00000197530	MIB2	NaN	chr1:[1564691,1564763)/.	NaN	0	0.5	
TH238	1565888-1565889	ENSG00000197530	MIB2	NaN	chr1:[1564946,1565017)/.	NaN	0	0.5	

```
df3c$VarIND <- paste0(df3c$var, "<->", df3c$gene)

pal = brewer.pal(7, "RdYlBu")

df3c$n_bin <- cut(df3c$n_sig, breaks = c(0, 1, 2, 3, Inf), right = F)

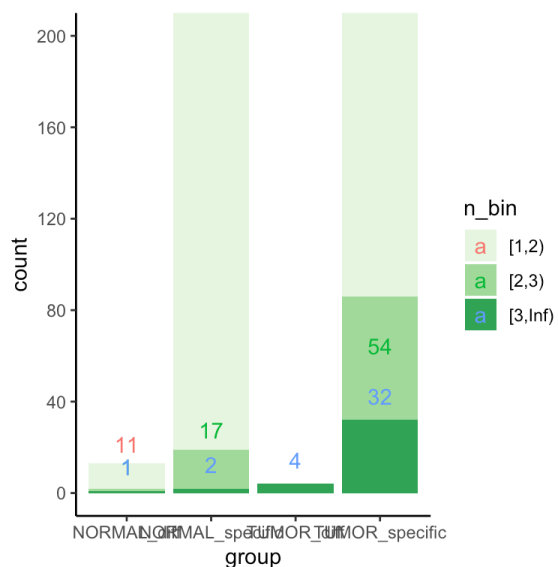
g1 <- ggplot(df3c, aes(x = group)) + geom_bar(aes(fill = n_bin)) + geom_text(stat = "count", aes(label = ..count..,
  color = n_bin), vjust = -1) + scale_fill_brewer(palette = "BuYlRd") + coord_cartesian(ylim = c(0,
  200)) + scale_y_continuous(breaks = c(0, 40, 80, 120, 160, 200)) + theme_classic()

g2 <- ggplot(df3c, aes(x = group)) + geom_bar(aes(fill = n_bin)) + geom_text(stat = "count", aes(label = ..count..,
  color = n_bin), vjust = -1) + scale_fill_brewer(palette = "BuYlRd") + coord_cartesian(ylim = c(200,
  3400)) + scale_y_continuous(breaks = c(200, 1000, 2000, 3000)) + theme_classic()

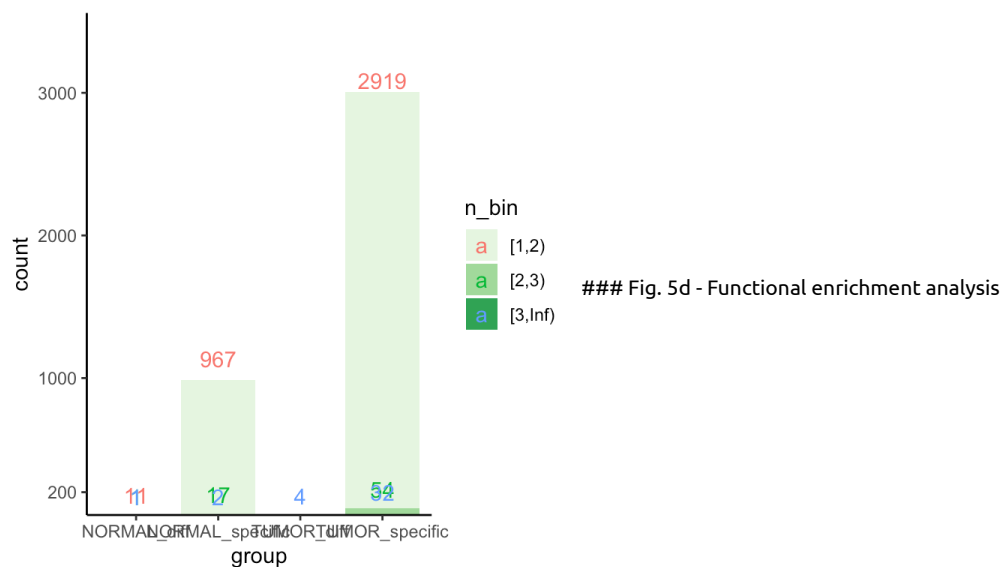
pdf("/Users/giovanni/hoffman_folder/micro_indel_project/FIGS/5b.pdf")
print(g1)
print(g2)
dev.off()
```

```
## quartz_off_screen
##                2
```

g1



g2



```
df3c <- read.csv('Cytoscape_terms.csv', sep = ',', header = T)
head(df3c)
```

Group	X.background.genes	X.genes	category	chart.color	description	FDR.value	genes	network.SUID	no
CANCER_SPECIFIC	101	2	Reactome Pathways	NA	ABC-family proteins mediated transport	0.03340	UBA52 UBC	2144	247
CANCER_SPECIFIC	31	2	Reactome Pathways	NA	Activated NOTCH1 Transmits Signal to the Nucleus	0.00720	UBA52 UBC	2144	247
CANCER_SPECIFIC	23	2	Reactome Pathways	NA	Activated tak1 mediates p38 mapk activation	0.00510	UBA52 UBC	2144	247
CANCER_SPECIFIC	17	3	Reactome Pathways	NA	Activation of IRF3/IRF7 mediated by TBK1/IKK epsilon	0.00037	UBA52 UBC IRF3	2144	247
CANCER_SPECIFIC	65	2	Reactome Pathways	NA	Activation of NF-kappaB in B cells	0.01850	UBA52 UBC	2144	247
NORMAL_DIFF	317	2	GO Biological Process	NA	Adaptive immune response	0.04150	HLA-E HLA-A	6009	620

```
df3c <- df3c[df3c$category %in% c("Reactome Pathways", "KEGG Pathways", "WikiPathways", "GO Biological Processes"),]
df3c <- df3c %>%
  arrange(desc(-log10(FDR.value))) %>%
  group_by(Group) %>%
  slice(1:3)

df3c$index <- 1:nrow(df3c)
df3c$descriptionx <- paste(df3c$description, df3c$index)
df3c$descriptionx <- factor(df3c$descriptionx, levels = reorder(df3c$descriptionx, -log10(df3c$FDR.value)))
df3c <- df3c[order(df3c$descriptionx),]

g1 <- ggplot(df3c, aes(y = -log10(FDR.value), x = descriptionx))+
  geom_bar(stat = 'identity', aes(fill = category)) +
  geom_text(aes(label = X..genes))+
  facet_grid(Group~., scales= 'free')+
  coord_flip()+
  theme_minimal()

pdf('/Users/giovanni/hoffman_folder/micro_indel_project/FIGS/5d.pdf', width = 14, height = 3)
print(g1)
dev.off()
```

```
## quartz_off_screen
##                2
```

```
g1
```

