BMI Final Project Supplementary File

Five Guys

Contents

Workflow Diagram

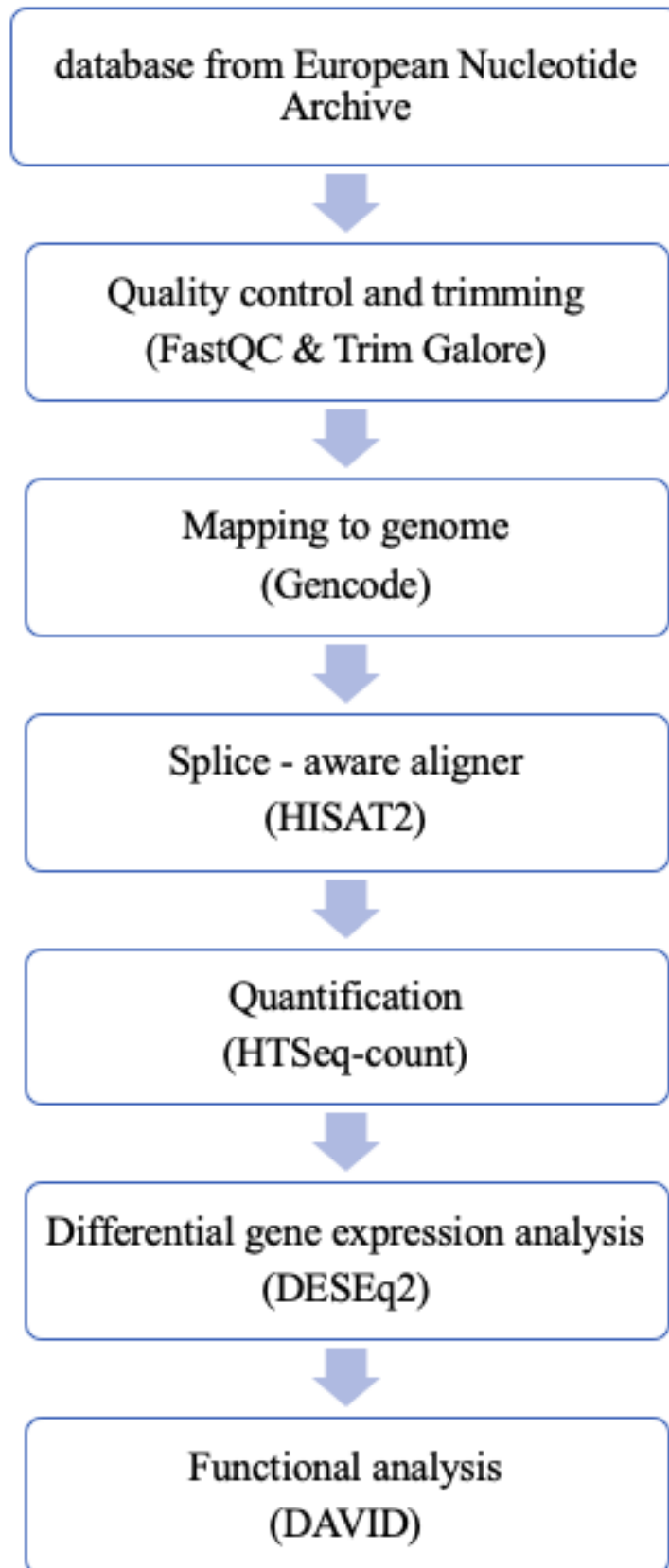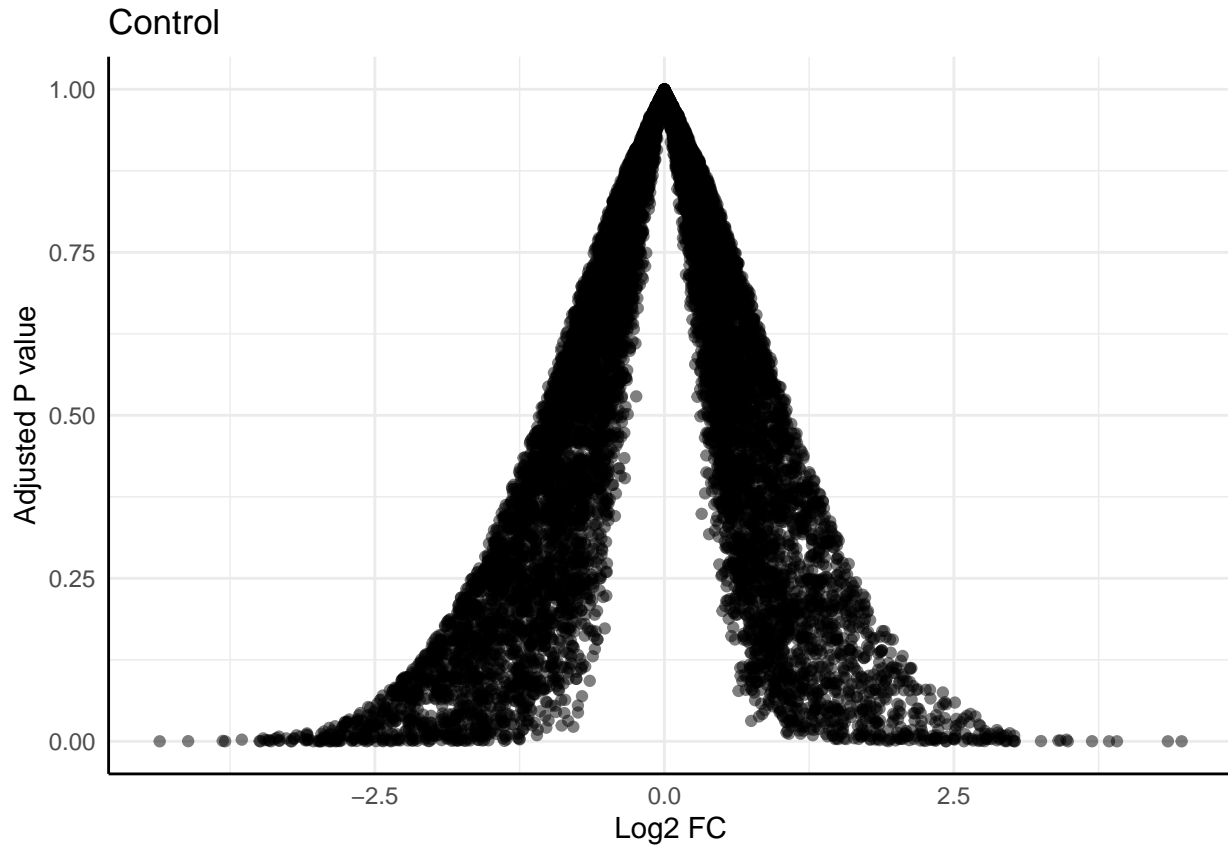Please see methods section for detailed overview of the project workflow.

Figure 1: Project workflow

Cases vs Controls R

Volcano Plot

```r
ggplot(data=control, aes(x=log2fc, y=padj)) +
  geom_jitter(alpha=0.5)+
  ggtitle("Control")+
  xlab("Log2 FC")+
  ylab("Adjusted P value")+
  ggsave("case_v_con_volcano.png", width=10, height=6)
```


Up vs Down Regulated Plot

```r
tab_con<-control%>%
  mutate(newdiff = ifelse((log2fc > 2 & padj < 0.05), "Up",
                      ifelse((log2fc < -2 & padj < 0.05), "Down", "lowdiff")))%>%
  mutate(abslog2fc = abs(log2fc))%>%
  filter(newdiff!="lowdiff")%>%
  group_by(newdiff)%>%
  slice_max(n=5, order_by=abslog2fc)%>%
  select(Position=X1, geneID, idlabel, log2fc, padj, direction=newdiff)

control%>%
  mutate(newdiff = ifelse((log2fc > 2 & padj < 0.05), "Up",
                      ifelse((log2fc < -2 & padj < 0.05), "Down", "Low-diff")))%>%
  mutate(newidlabel=ifelse(idlabel %in% tab_con$idlabel, idlabel, NA))%>%
```

```r
ggplot(aes(x=log2fc, y=-log10(padj), col=newdiff, label=newidlabel)) +
geom_point(aes(alpha=log10(padj)), show.legend=F) +
geom_vline(xintercept=c(-2, 2), col="black", linetype = "dashed") +
geom_hline(yintercept=-log10(0.05), col="black", linetype = "dashed")+
annotate("label", x = -3.5, y = 0, label = "Down Regulated", color="#ba3e14",fontface = "bold")+
annotate("label", x = 4, y = 0, label = "Up Regulated", color="#1333b0",fontface = "bold")+
ggrepel::geom_label_repel(min.segment.length = unit(0, 'lines'), show.legend=F)+
scale_color_manual(values=c("#ba3e14", "#b8bfb8", "#1333b0"))+
labs(title="Differential Gene Expression", subtitle="5 Control Samples and 5 Melanoma Samples")+
ylab("Negative Log Adjusted P value")+
xlab("Log base 2 FC")+
ylim(0,10)+
theme_minimal()
```
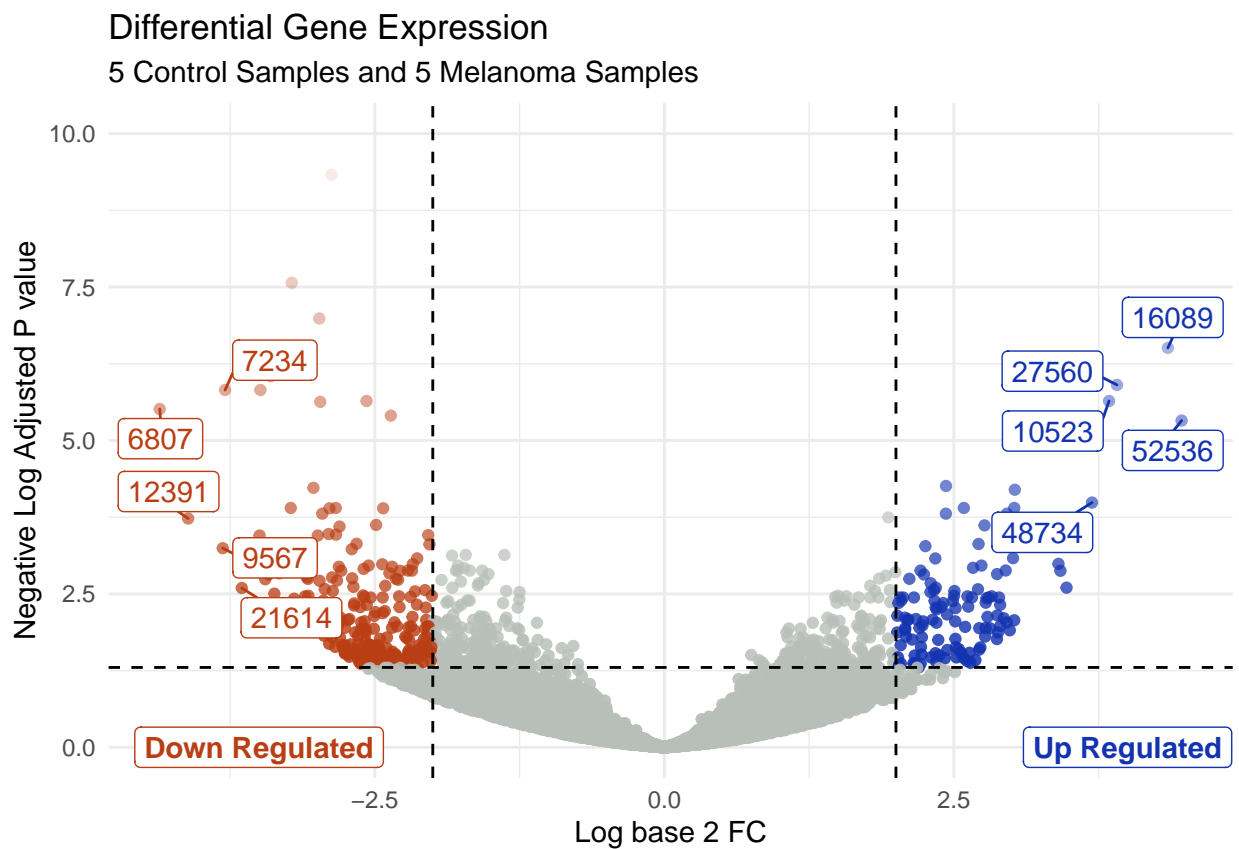


```r
ggsave("case_v_con.png", width=10, height=6)
```

Tabular Views

```r
tab_con<-control%>%
  mutate(newdiff = ifelse((log2fc > 2 & padj < 0.05), "Up",
                          ifelse((log2fc < -2 & padj < 0.05), "Down", "lowdiff")))%>%
  mutate(abslog2fc = abs(log2fc))%>%
  filter(newdiff!="lowdiff")%>%
  group_by(newdiff)%>%
  slice_max(n=5, order_by=abslog2fc)%>%
  select(Position=X1, geneID, idlabel, log2fc, padj, direction=newdiff)
```

```
tab_con%>%
  filter(direction=="Down")
```

Position	geneID	idlabel	log2fc	padj	direction
12	ENSG00000133872.14	6807	-4.358153	0.0000031	Down
30	ENSG00000168497.5	12391	-4.112331	0.0001870	Down
45	ENSG00000152558.15	9567	-3.813681	0.0005698	Down
8	ENSG00000136156.15	7234	-3.793525	0.0000015	Down
109	ENSG00000210194.1	21614	-3.649254	0.0025375	Down

```
tab_con%>%
  filter(direction=="Up")
```

Position	geneID	idlabel	log2fc	padj	direction
14	ENSG00000270090.5	52536	4.467033	0.0000047	Up
4	ENSG00000185818.8	16089	4.348400	0.0000003	Up
6	ENSG00000226869.6	27560	3.908796	0.0000012	Up
9	ENSG00000160678.12	10523	3.839950	0.0000023	Up
19	ENSG00000261857.7	48734	3.693108	0.0001029	Up

```
control%>%
  mutate(newdiff = ifelse((log2fc > 2 & padj < 0.05), "Up",
                    ifelse((log2fc < -2 & padj < 0.05), "Down", "lowdiff")))%>%
  mutate(abslog2fc = abs(log2fc))%>%
  dplyr::group_by(newdiff)%>%
  dplyr::summarize(count=n())%>%
  tidyr::pivot_wider(names_from=newdiff, values_from=count)%>%
  select(Down, `No Significant Change`=lowdiff, Up)
```

Down	No Significant Change	Up
236	11845	119

Good vs Bad Prognosis R

Volcano Plot

```
ggplot(data=cases, aes(x=log2fc, y=padj)) +
  geom_jitter(alpha=0.5)+
  ggtitle("Cases")+
  xlab("Log2 FC")+
  ylab("Adjusted P value")+
  ggsave("bad_v_good_volcano.png", width=10, height=6)
```
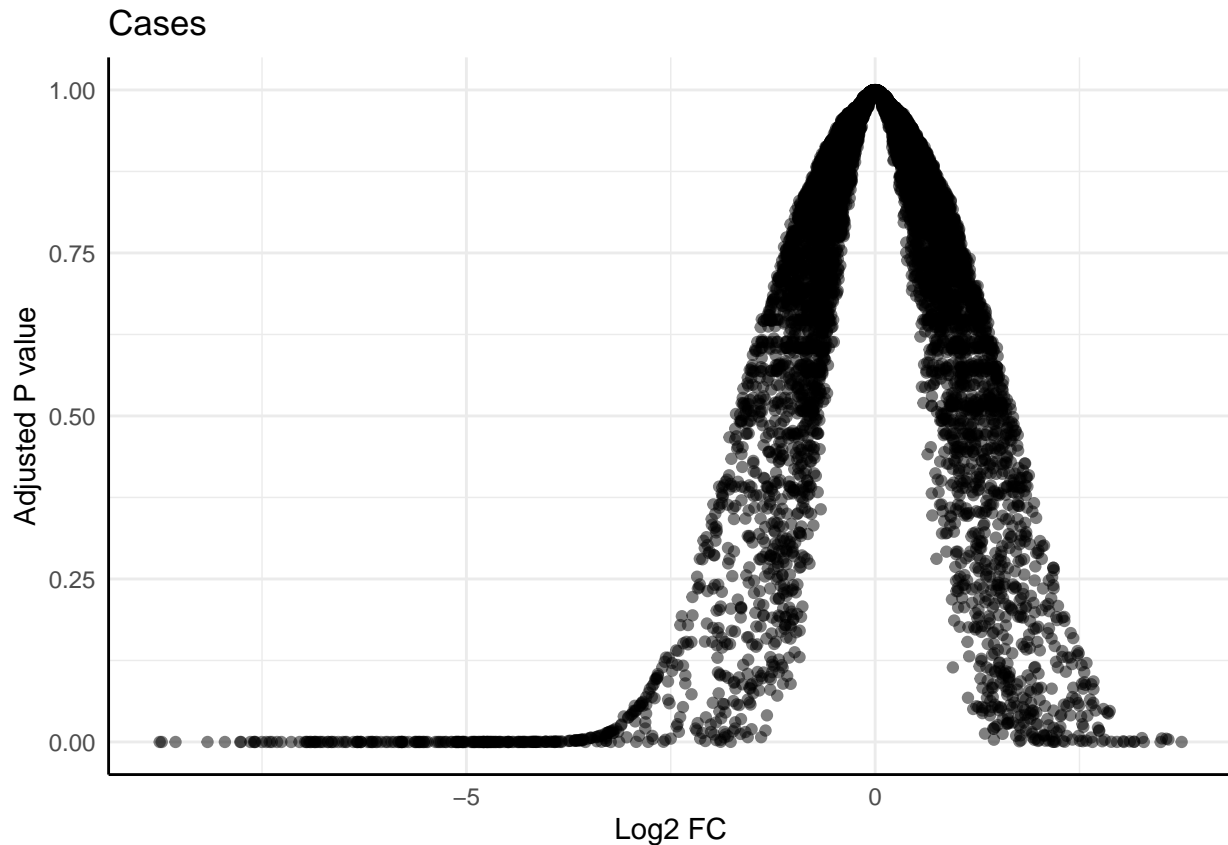
Cases

Up vs Down Regulated Plot

```r
tab_case<-cases%>%
  mutate(newdiff = ifelse((log2fc > 2 & padj < 0.05), "Up",
                          ifelse((log2fc < -2 & padj < 0.05), "Down", "lowdiff")))%>%
  mutate(abslog2fc = abs(log2fc))%>%
  filter(newdiff!="lowdiff")%>%
  group_by(newdiff)%>%
  slice_max(n=5, order_by=abslog2fc)%>%
  select(Position=X1, geneID, idlabel, log2fc, padj, direction=newdiff)

cases%>%
  mutate(newdiff = ifelse((log2fc > 2 & padj < 0.05), "Up",
                          ifelse((log2fc < -2 & padj < 0.05), "Down", "Low-diff")))%>%
  mutate(newidlabel=ifelse(idlabel %in% tab_case$idlabel, idlabel, NA))%>%
  ggplot(aes(x=log2fc, y=-log10(padj), col=newdiff, label=newidlabel)) +
  geom_point(aes(alpha=log10(padj)), show.legend=F) +
  geom_vline(xintercept=c(-2, 2), col="black", linetype = "dashed") +
  geom_hline(yintercept=-log10(0.05), col="black", linetype = "dashed")+
  annotate("label", x = -3.5, y = 40, label = "Down Regulated", color="#ba3e14", fontface = "bold")+
  annotate("label", x = 3.25, y = 40, label = "Up Regulated", color="#1333b0", fontface = "bold")+
  ggrepel::geom_label_repel(min.segment.length = unit(0, 'lines'), box.padding=0.65, show.legend=F)+
  scale_color_manual(values=c("#ba3e14", "#b8bfb8", "#1333b0"))+
  labs(title="Differential Gene Expression in Melanoma Patients", subtitle="Bad Prognosis vs Good Progno
  ylab("Negative Log Adjusted P value")+
  xlab("Log base 2 FC")+
```

```
ylim(0,40)+
theme_minimal()+
ggsave("bad_v_good.png", width=10, height=6)
```

Differential Gene Expression in Melanoma Patients
Bad Prognosis vs Good Prognosis
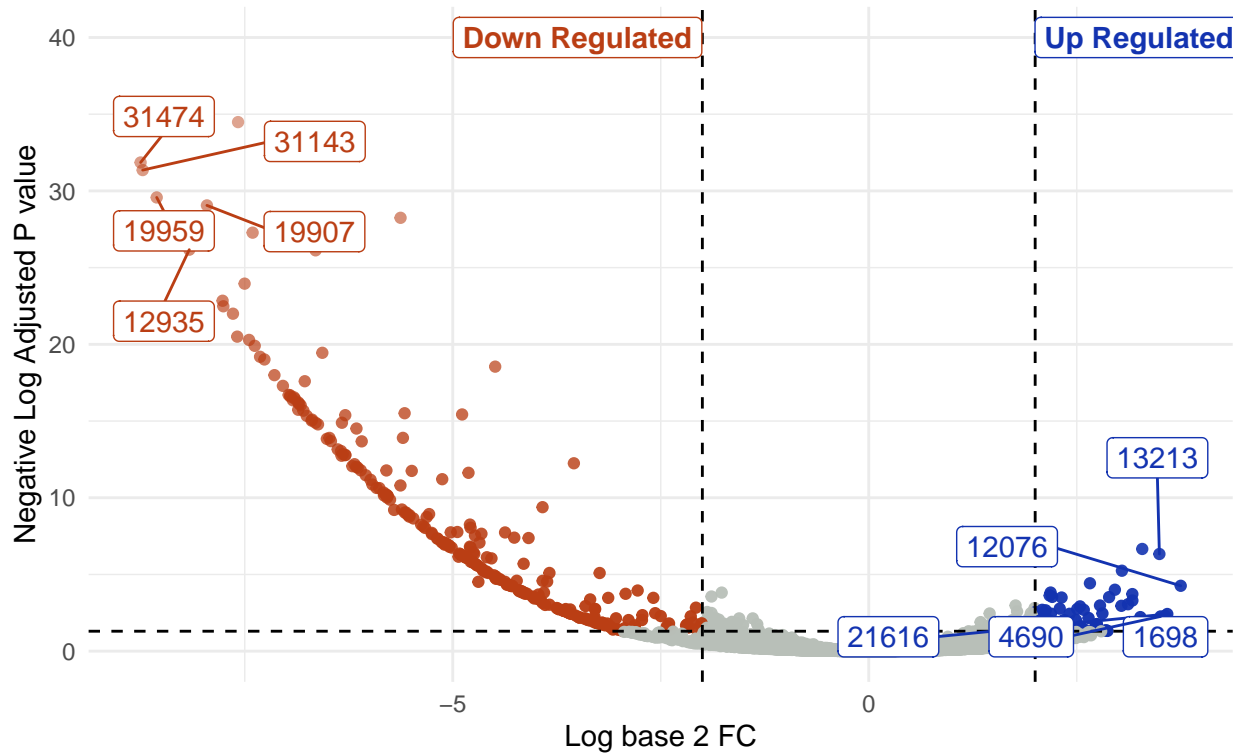
Tabular Views

```
tab_case%>%
  filter(direction=="Down")
```

Position	geneID	idlabel	log2fc	padj	direction
3	ENSG00000231925.12	31474	-8.751542	0	Down
4	ENSG00000231500.7	31143	-8.726711	0	Down
5	ENSG00000204592.9	19959	-8.558053	0	Down
9	ENSG00000170889.14	12935	-8.166531	0	Down
6	ENSG00000204469.13	19907	-7.954339	0	Down

```
tab_case%>%
  filter(direction=="Up")
```

Position	geneID	idlabel	log2fc	padj	direction
167	ENSG00000167286.10	12076	3.749711	0.0000548	Up
248	ENSG00000210196.2	21616	3.587957	0.0037827	Up
251	ENSG00000086548.9	1698	3.561389	0.0049815	Up
255	ENSG00000117228.10	4690	3.508025	0.0054362	Up

7

Position	geneID	idlabel	log2fc	padj	direction
125	ENSG00000172005.11	13213	3.493104	0.0000005	Up

```r
cases%>%
  mutate(newdiff = ifelse((log2fc > 2 & padj < 0.05), "Up",
                     ifelse((log2fc < -2 & padj < 0.05), "Down", "lowdiff")))%>%
  mutate(abslog2fc = abs(log2fc))%>%
  dplyr::group_by(newdiff)%>%
  dplyr::summarize(count=n())%>%
  tidyr::pivot_wider(names_from=newdiff, values_from=count)%>%
  select(Down, `No Significant Change`=lowdiff, Up)
```

Down	No Significant Change	Up
267	9872	59