Final project

Annie Cohen

3/26/2022

```r
cancerdata <- data %>%
  mutate(prop_infiltrated = infiltration_count/tile_count)

long_data <- cancerdata %>%
  pivot_longer(-colnames(cancerdata[,c(1:8,34)]), names_to = "gene", values_to = "exp")

genes <- c("TLR4","CIITA","KLRD1","MC3R","NAMPT")

## prop infiltrated ~ gene expression * gender

long_data %>%
  filter(gene %in% genes) %>%
  ggplot(aes(x = exp, y = prop_infiltrated, col = gender)) +
  geom_point(alpha = 0.5) +
  geom_smooth(method = 'lm', formula = 'y~x', se = FALSE) +
  facet_wrap(~gene) +
  scale_color_manual(values = c("cornflowerblue","darkgoldenrod2")) +
  theme_classic()
```
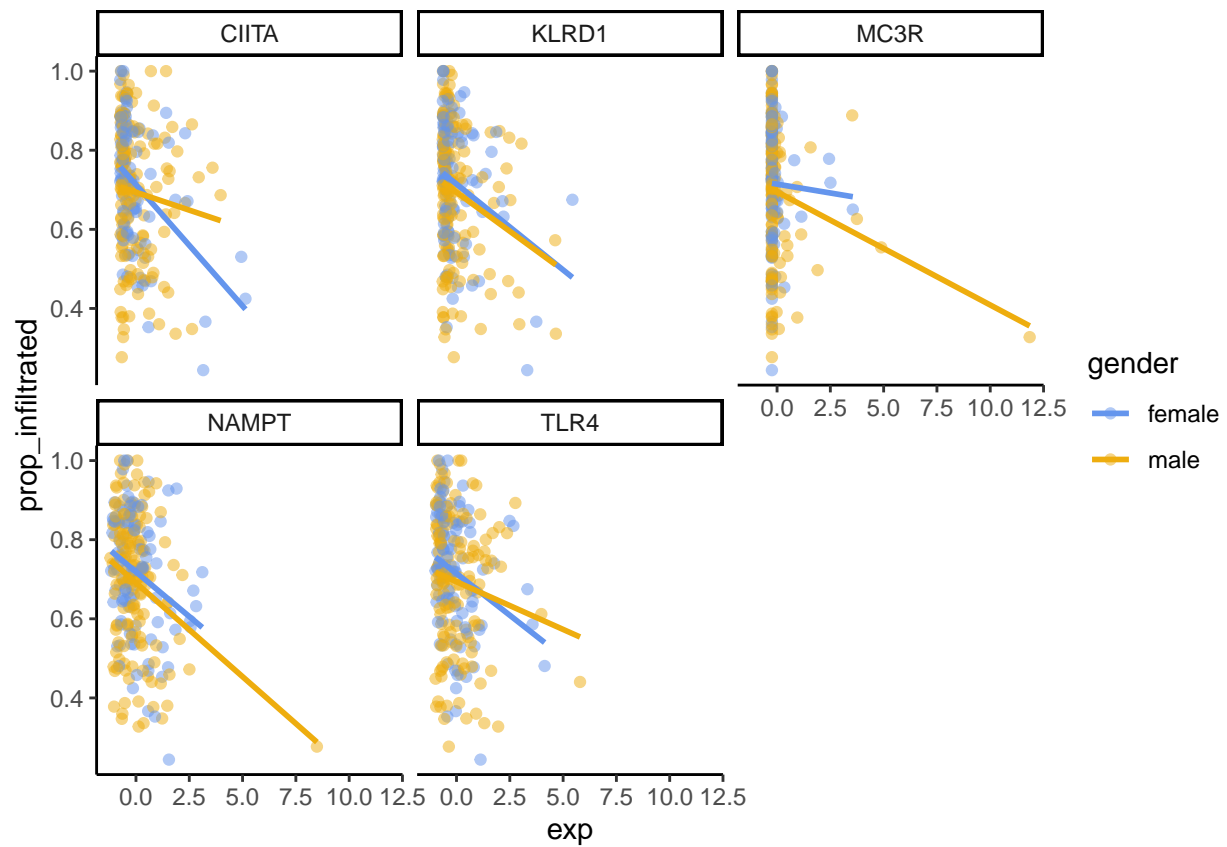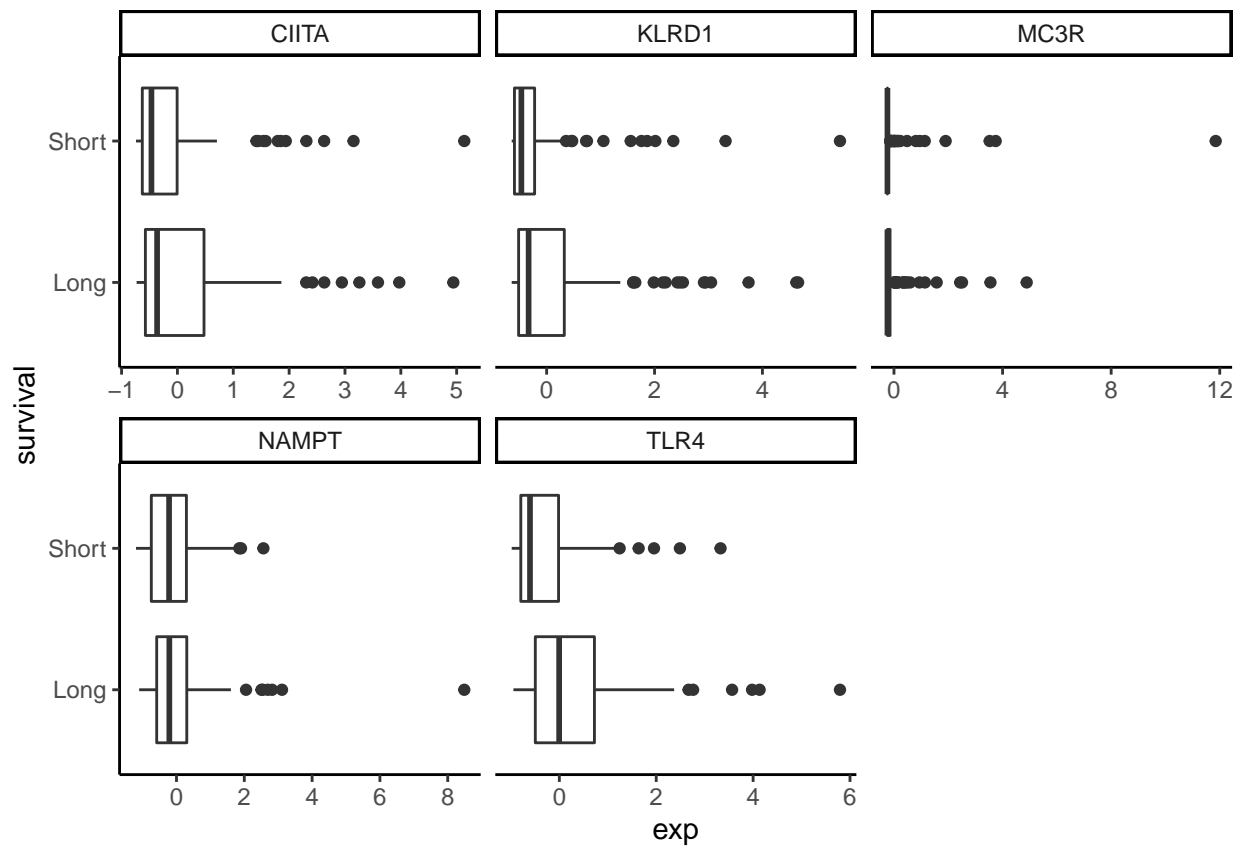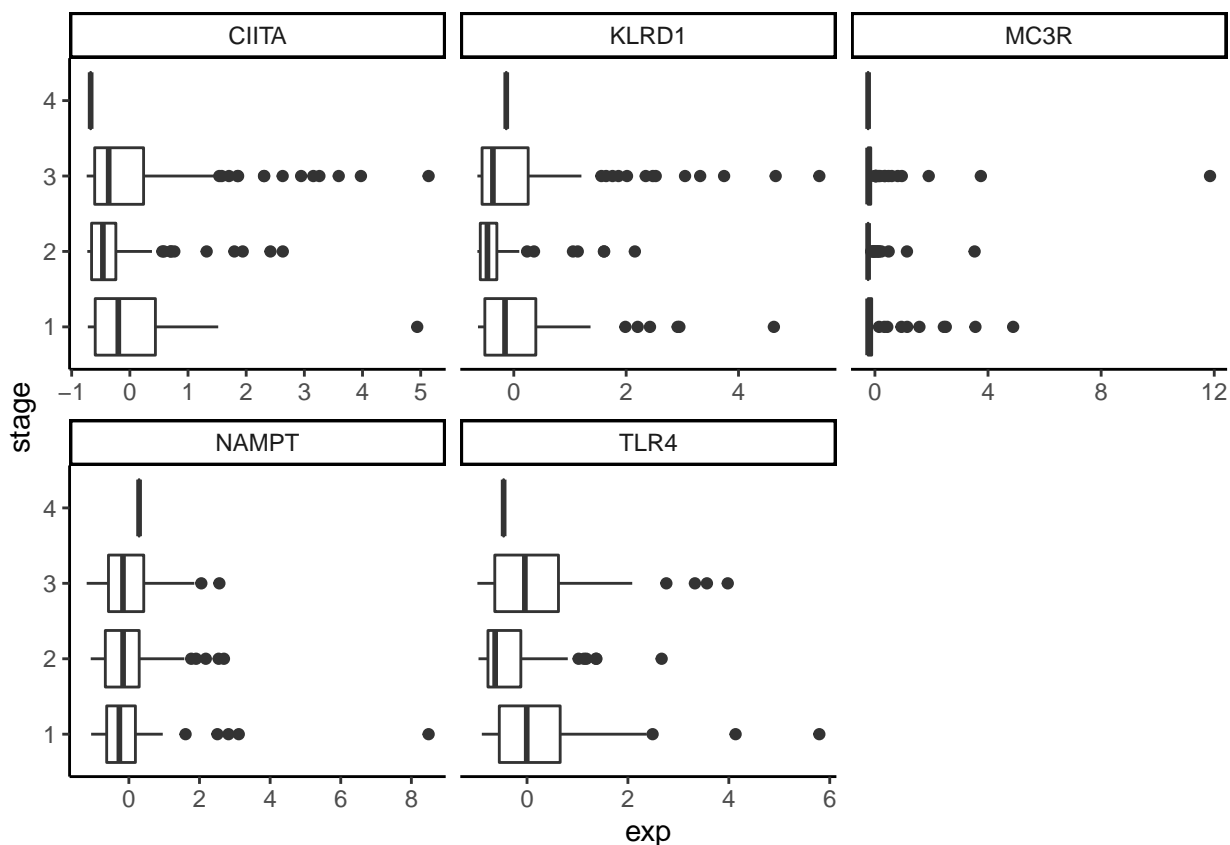
```
## survival ~ gene expression

long_data %>%
  filter(gene %in% genes & !is.na(survival)) %>%
  ggplot(aes(x = exp, y = survival)) +
  geom_boxplot() +
  facet_wrap(~gene, scales = "free_x") +
  theme_classic()
```

```
## stage ~ gene expression

long_data %>%
  filter(gene %in% genes & !is.na(stage)) %>%
  ggplot(aes(x = exp, y = stage)) +
  geom_boxplot(aes(group = stage)) +
  facet_wrap(~gene, scales = "free_x") +
  theme_classic()
```

```
## tbl_summary object

cancerdata %>%
  tbl_summary(include = -id, by = survival) %>%
  add_p()
```

```
## 4 observations missing `survival` have been removed. To include these observations, use `forcats::fc
```

```
## Table printed with `knitr::kable()`, not {gt}. Learn why at
## https://www.danieldsjoberg.com/gtsummary/articles/rmarkdown.html
## To suppress this message, include `message = FALSE` in code chunk header.
```

Characteristic	Long, N = 127	Short, N = 109	p-value
stage			<0.001
1	47 (37%)	7 (6.4%)	
2	36 (28%)	38 (35%)	
3	44 (35%)	63 (58%)	
4	0 (0%)	1 (0.9%)	
gender			0.2
female	45 (35%)	48 (44%)	
male	82 (65%)	61 (56%)	
age	54 (43, 67)	61 (51, 74)	0.001
tile_count	322 (214, 462)	308 (165, 477)	0.5
infiltration_count	197 (131, 284)	218 (123, 312)	0.7

Characteristic	**Long**, N = 127	**Short**, N = 109	**p-value**
trans_class			<0.001
immune	80 (63%)	37 (34%)	
keratin	21 (17%)	55 (50%)	
MITF-low	26 (20%)	17 (16%)	
AKT1	-0.38 (-0.70, 0.04)	0.08 (-0.58, 0.76)	<0.001
AKT3	0.08 (-0.56, 0.73)	-0.47 (-0.96, 0.38)	<0.001
DCK	0.12 (-0.46, 0.67)	-0.33 (-0.93, 0.32)	<0.001
INSR	0.07 (-0.42, 0.74)	-0.35 (-0.80, 0.24)	<0.001
MC1R	-0.44 (-0.54, -0.13)	-0.15 (-0.48, 0.50)	<0.001
PAK2	0.18 (-0.25, 0.84)	-0.43 (-1.03, 0.36)	<0.001
TLR4	0.00 (-0.50, 0.72)	-0.61 (-0.80, -0.02)	<0.001
CIITA	-0.36 (-0.57, 0.48)	-0.47 (-0.63, -0.01)	0.013
IL17RB	-0.23 (-0.43, 0.22)	-0.26 (-0.56, 0.05)	0.2
KLRD1	-0.33 (-0.52, 0.33)	-0.47 (-0.60, -0.22)	0.001
MC3R	-0.24 (-0.24, -0.14)	-0.24 (-0.24, -0.24)	0.7
NAMPT	-0.21 (-0.58, 0.30)	-0.22 (-0.74, 0.29)	0.3
RFX5	-0.09 (-0.60, 0.75)	-0.36 (-0.86, 0.25)	0.012
SOS1	-0.04 (-0.44, 0.40)	-0.33 (-0.71, 0.14)	0.002
DDX58	-0.16 (-0.46, 0.47)	-0.43 (-0.68, 0.05)	<0.001
EIF2AK2	-0.09 (-0.57, 0.55)	-0.35 (-0.81, 0.17)	0.002
APOBEC3G	-0.13 (-0.58, 0.60)	-0.43 (-0.71, -0.07)	<0.001
IL15RA	-0.21 (-0.59, 0.58)	-0.50 (-0.72, -0.02)	0.001
CCL28	-0.19 (-0.76, 0.49)	-0.17 (-0.65, 0.32)	0.7
IL17C	-0.21 (-0.45, 0.06)	-0.45 (-0.45, -0.01)	0.14
B2M	-0.04 (-0.62, 0.70)	-0.48 (-0.84, 0.07)	0.006
CD86	-0.12 (-0.55, 0.76)	-0.58 (-0.78, -0.05)	<0.001
CD81	-0.40 (-0.72, 0.03)	-0.09 (-0.39, 0.52)	<0.001
C5	-0.17 (-0.48, 0.32)	-0.38 (-0.68, 0.11)	0.001
TNFRSF19	-0.32 (-0.60, 0.28)	-0.40 (-0.63, 0.21)	0.2
prop_infiltrated	0.69 (0.56, 0.79)	0.77 (0.64, 0.86)	0.002

```
colSums(is.na(cancerdata)) %>%
  kable()
```

	x
id	0
stage	0
survival	4
gender	0
age	0
tile_count	0
infiltration_count	0
trans_class	0
AKT1	0
AKT3	0
DCK	0
INSR	0
MC1R	0
PAK2	0
TLR4	0

	x
CIITA	0
IL17RB	0
KLRD1	0
MC3R	0
NAMPT	0
RFX5	0
SOS1	0
DDX58	0
EIF2AK2	0
APOBEC3G	0
IL15RA	0
CCL28	0
IL17C	0
B2M	0
CD86	0
CD81	0
C5	0
TNFRSF19	0
prop_infiltrated	0