Plotting prediction outputs

Initialise environment

```
library(ggplot2)
library(dplyr)
```

```
##
## Attaching package: 'dplyr'
```

```
## The following objects are masked from 'package:stats':
##
##     filter, lag
```

```
## The following objects are masked from 'package:base':
##
##     intersect, setdiff, setequal, union
```

```
library(stringr)

theme_sv <-   theme_bw(base_size = 7, base_family = "Helvetica") +
  theme(strip.background = element_blank(),
        axis.text.y = element_text(color = "black"),
        axis.text.x = element_text(color = "black"),
        text = element_text(family = "Helvetica"),
        legend.text=element_text(size=7),
        legend.key.size = unit(0.25, "cm"),
        panel.grid = element_blank(),
        panel.border = element_rect(color = "black"))
```

Read in data and process

```
otx2_scores <- read.csv('data/ground_truth.tsv', sep='\t', row.names = 1)
predictions <- read.csv('data/otx2_predictions_in_tissues.csv', row.names = 1)
all_scores <- cbind(otx2_scores, predictions[rownames(otx2_scores),])

pred_corrs <- c()
for(i in 1:ncol(predictions)){
  pred_corrs <- c(pred_corrs,
                  cor(all_scores$meanexpressionscoreavg,
                      all_scores[, 1 + i]))
}
pred_corrs <- data.frame(pred_corrs)
colnames(pred_corrs) <- 'Pearson.r'
pred_corrs$tissue <- colnames(all_scores[,-c(1)])
```

Plot predictions

```
mean(pred_corrs$Pearson.r)
```

```
## [1] -0.1613937
```

```
max(pred_corrs$Pearson.r)
```

```
## [1] 0.1605401
```

```
pred_corrs[order(pred_corrs$Pearson.r, decreasing = TRUE),]
```

```
##       Pearson.r               tissue
## 29  0.16054006          RNA.thyroid
## 14  0.13821309           RNA.kidney
## 21  0.07713746         RNA.pituitary
## 18  0.02331026            RNA.nerve
## 16 -0.02225651             RNA.lung
## 7  -0.02824873            RNA.brain
## 23 -0.04629775  RNA.salivary_gland
## 8  -0.05401382           RNA.breast
## 20 -0.05627263         RNA.pancreas
## 17 -0.07230418           RNA.muscle
## 12 -0.08190261  RNA.fallopian_tube
## 15 -0.10267514            RNA.liver
## 13 -0.10508633            RNA.heart
## 1  -0.13878299  RNA.adipose_tissue
## 2  -0.18461238   RNA.adrenal_gland
## 9  -0.18566125     RNA.cervix_uteri
## 22 -0.18840678         RNA.prostate
## 5  -0.19498981     RNA.blood_vessel
## 19 -0.20263888            RNA.ovary
## 27 -0.20452899          RNA.stomach
## 4  -0.23273535            RNA.blood
## 30 -0.25272661           RNA.uterus
## 24 -0.26358479             RNA.skin
## 3  -0.27572194          RNA.bladder
## 11 -0.28039373        RNA.esophagus
## 31 -0.32302796           RNA.vagina
## 28 -0.33130668           RNA.testis
## 26 -0.35370417           RNA.spleen
## 25 -0.39616962 RNA.small_intestine
## 6  -0.40114510      RNA.bone_marrow
## 10 -0.42321095            RNA.colon
```

```
ggplot(pred_corrs, aes(x=Pearson.r)) +
  geom_histogram(bins=40) +
  labs(x='Pearson r', y='Count') +
  theme_sv +
  scale_x_continuous(limits=c(-1,1))
```
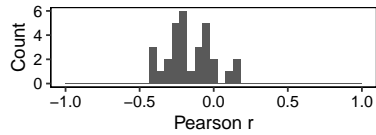
```
data.frame(blood=all_scores$RNA.blood,
           true=all_scores$meanexpressionscoreavg,
           wt_marker=ifelse(rownames(all_scores) == 'wt', 'WT', 'NA')) %>%
  ggplot(aes(y=blood, x=true, shape=wt_marker)) +
  geom_point(size=.5) +
  theme_sv +
  labs(x='Relative gene expression',
       y='Predicted OTX2 RNA-seq\ncoverage [GTEx:blood]') +
  guides(shape='none')
```