

# Correlation Matrix & Association Signals

Brian Jo Hsuan Lee

5/6/2022

Load data and acquire the correlation matrix without non-gene columns. Save the correlation matrix.

```
data =  
  read_csv("~/Desktop/Columbia/Spring_2022/P8139-Statistical_Genetic_Modeling/Project/BRCA_data") %>%  
  arrange(id) %>%  
  mutate(data = factor(data), id = factor(id))  
  
gene_cor =  
  data %>%  
  select(-data, -id) %>%  
  cor() %>%  
  data.frame()  
  
gene_cor[is.na(gene_cor)] = 0  
  
write_csv(gene_cor, file="/Users/beelee/Desktop/Columbia/Spring_2022/P8139-Statistical_Genetic_Modeling/Project/BRCA_data/gene_cor.csv")
```

Create a vector of adjusted p-values for running a paired t-test on each gene. Save the association significance.

```
p_vals =  
  data[1, ] %>%  
  select(-data, -id)  
  
for(i in seq_len(2000)){  
  paired_data =  
    data %>%  
    rename(this = names(data[, i+2])) %>%  
    select(data, id, this) %>%  
    filter(as.integer(id)>244) %>%  
    pivot_wider(  
      names_from = data,  
      values_from = this  
    )  
  test_obj = t.test(paired_data$case, paired_data$control, paired = TRUE, alternative = 'two.sided')  
  p_vals[1, i] = test_obj$p.value  
}  
  
corr_p_vals = p.adjust(p_vals[1,], method = "BH")  
  
p_vals_t =  
  data.frame(  
    'gene' = names(corr_p_vals),  
    'pval' = corr_p_vals  
  ) %>%
```

```
arrange(pval)

write_csv(p_vals_t, file="/Users/beelee/Desktop/Columbia/Spring_2022/P8139-Statistical_Genetic_Modeling
```