

Untitled

```
tpm_id <- "syn12385635"

tpm_df <- tpm_id %>%
  create_df_from_synapse_id %>%
  .[,order(colnames(.))]

tpm_m <- tpm_df %>%
  df_to_matrix("Hugo")

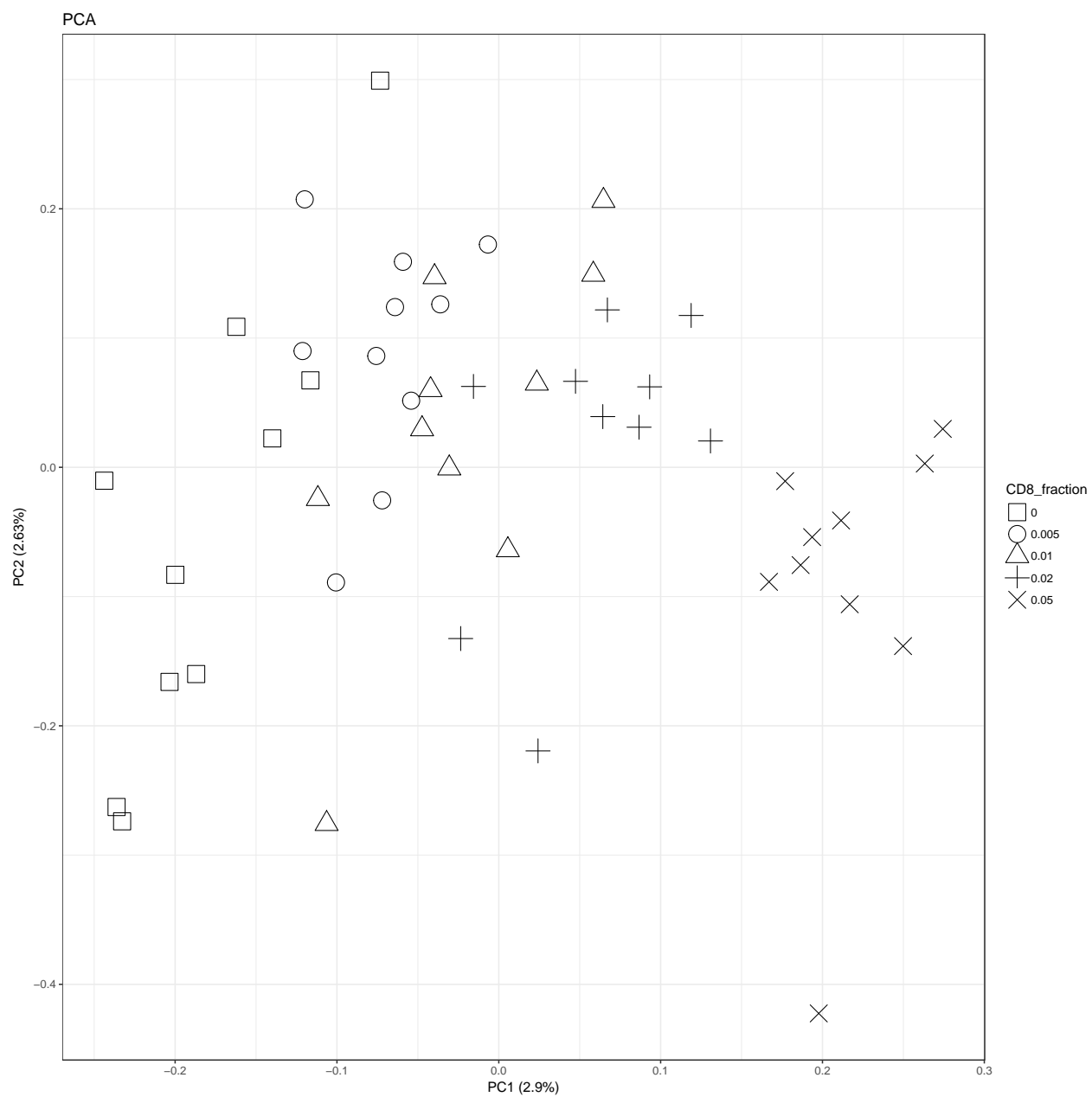
ltpm_df <- tpm_m %>%
  add(1) %>%
  log10 %>%
  matrix_to_df("Hugo")

ltpm_matrix_tranpose <- ltpm_df %>%
  df_to_matrix("Hugo") %>%
  t

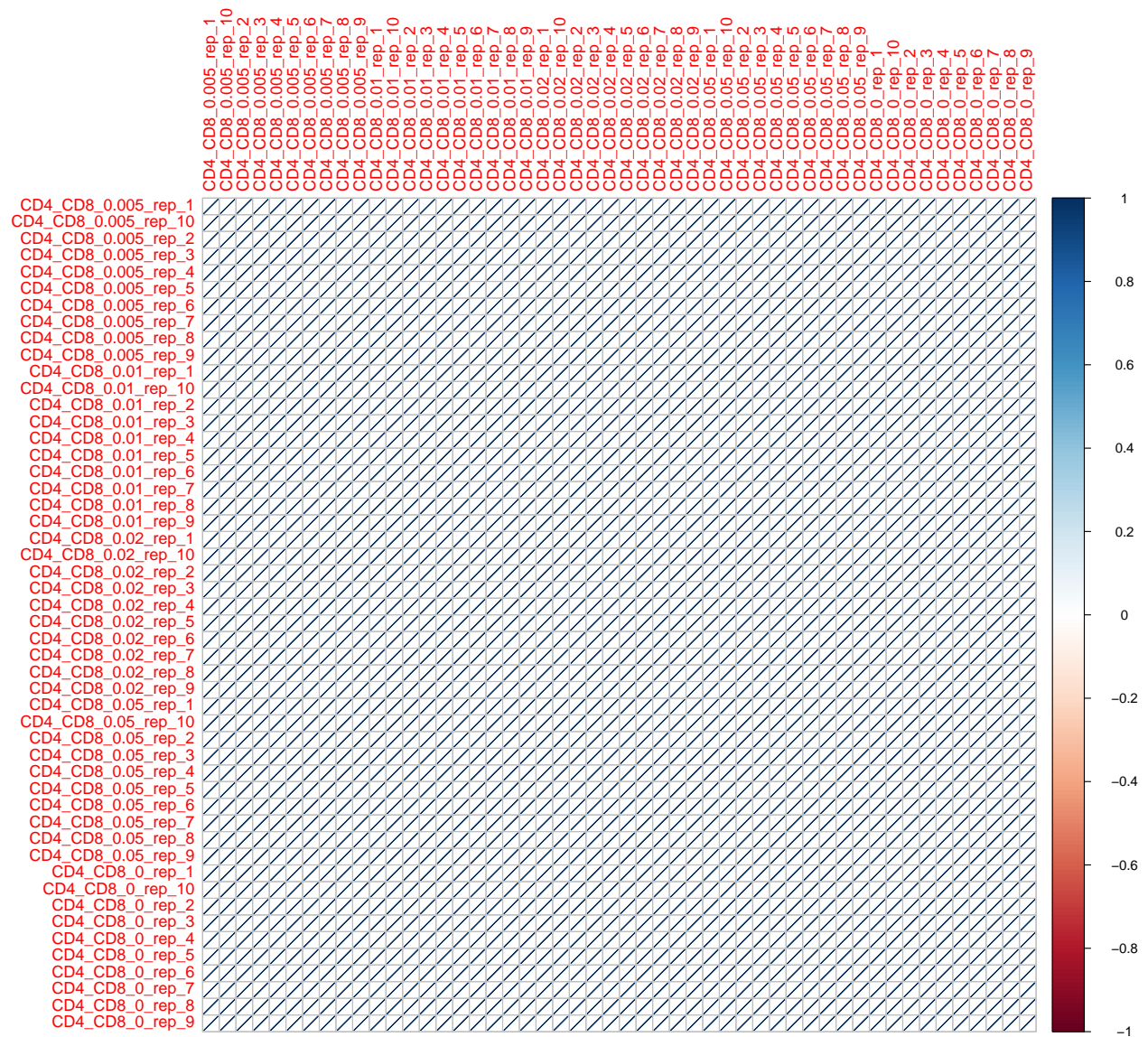
sample_df <- ltpm_matrix_tranpose %>%
  matrix_to_df("sample") %>%
  dplyr::select(sample) %>%
  mutate(CD8_fraction = str_match(sample, "CD4_CD8_([:print:]+)_rep_([:digit:]+")[,2]) %>%
  mutate(sample_n = str_match(sample, "CD4_CD8_([:print:]+)_rep_([:digit:]+")[,2])

annotation_df <- sample_df %>%
  as.data.frame %>%
  column_to_rownames("sample")

autoplot(
  prcomp(ltpm_matrix_tranpose),
  data = annotation_df,
  shape = "CD8_fraction",
  size = 6,
  main = "PCA") +
  scale_shape_manual(values = 0:6) +
  theme_bw()
```

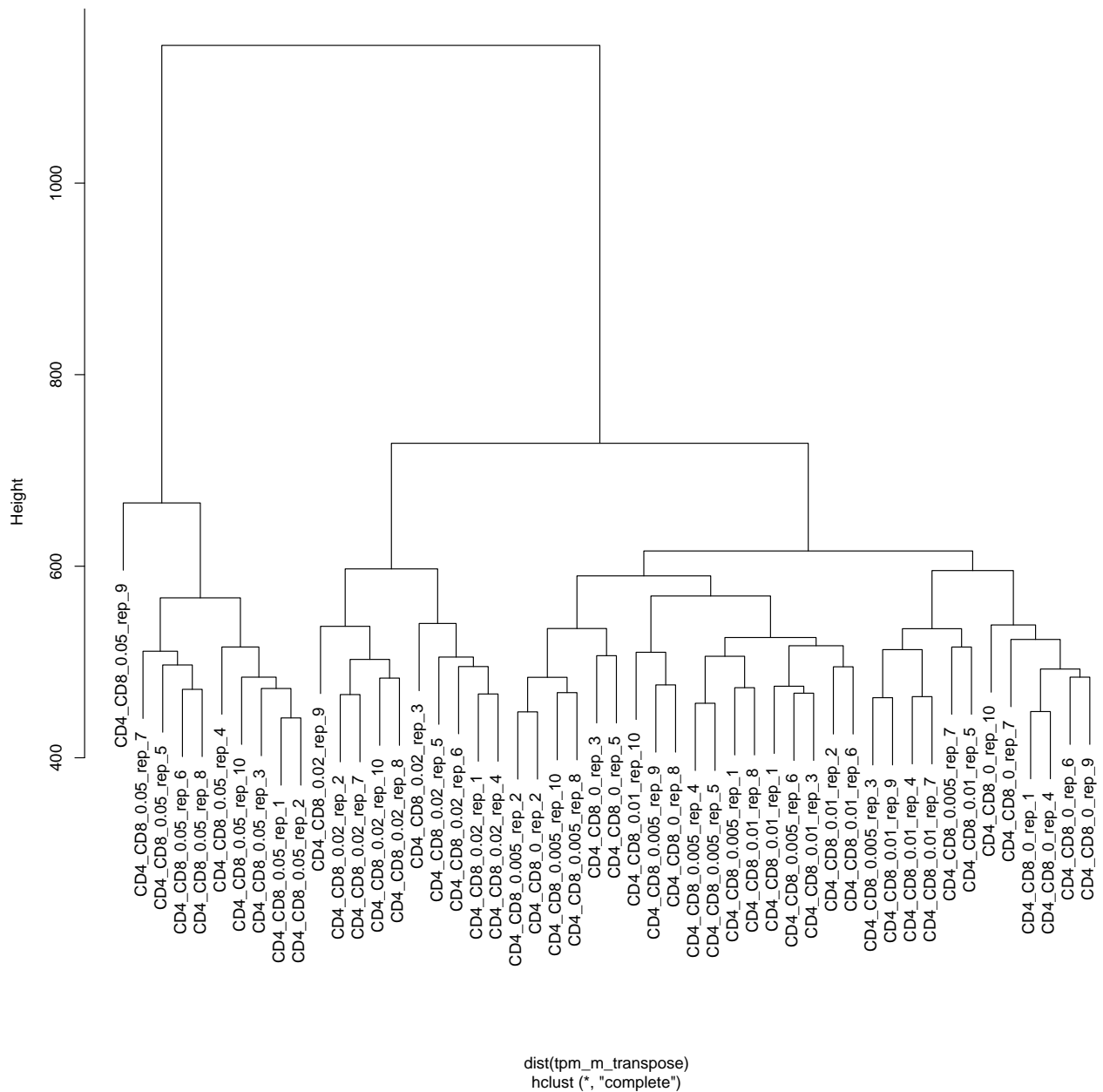


```
cor_matrix <- cor(tpm_m)
corrplot(cor_matrix, method = "ellipse")
```



```
tpm_m_transpose <- t(tpm_m)
clusters <- hclust(dist(tpm_m_transpose))
plot(clusters)
```

Cluster Dendrogram



```
gene_id <- "syn11918430"
cibersort_gene_id <- "syn12184137"

# mcp genes -----
mcp_genes <- gene_id %>%
  create_df_from_synapse_id %>%
  filter(Method == "mcpcounter") %>%
  split(.$cell_type) %>%
  map(use_series, Hugo)

mcp_ssgsea_df <- gsva(tpm_m, mcp_genes, rnaseq = FALSE, method = "ssgsea", verbose = F) %>%
  matrix_to_df("MCP_cell_type") %>%
```

```

gather(key = "sample", value = "enrichment", CD4_CD8_0.005_rep_1:CD4_CD8_0_rep_9) %>%
left_join(sample_df)

mcp_CD8_pval_df <- mcp_ssgsea_df %>%
  filter(MCP_cell_type == "CD8 T cells") %>%
  dplyr::select(sample_n, enrichment, CD8_fraction) %>%
  mutate(CD8_fraction = str_c("CD8_fraction_", CD8_fraction)) %>%
  spread(key = "CD8_fraction", value = enrichment) %>%
  dplyr::select(-sample_n)

mcp_CD8_pvals <- mcp_CD8_pval_df %>%
  colnames %>%
  .[-1] %>%
  map(function(col) wilcox.test(mcp_CD8_pval_df$CD8_fraction_0, extract2(mcp_CD8_pval_df , col))) %>%
  map_dbl(extract2, "p.value") %>%
  set_names(colnames(mcp_CD8_pval_df[-1]))

mcp_CL_pval_df <- mcp_ssgsea_df %>%
  filter(MCP_cell_type == "Cytotoxic lymphocytes") %>%
  dplyr::select(sample_n, enrichment, CD8_fraction) %>%
  mutate(CD8_fraction = str_c("CD8_fraction_", CD8_fraction)) %>%
  spread(key = "CD8_fraction", value = enrichment) %>%
  dplyr::select(-sample_n)

mcp_CL_pvals <- mcp_CL_pval_df %>%
  colnames %>%
  .[-1] %>%
  map(function(col) wilcox.test(mcp_CD8_pval_df$CD8_fraction_0, extract2(mcp_CL_pval_df , col))) %>%
  map_dbl(extract2, "p.value") %>%
  set_names(colnames(mcp_CL_pval_df[-1]))

# cibersort genes -----

cs_genes <- cibersort_gene_id %>%
  create_df_from_synapse_id %>%
  split(.$cell_type) %>%
  map(use_series, Hugo)

cs_ssgsea_df <- gsva(tpm_m, cs_genes, rnaseq = FALSE, method = "ssgsea", verbose = F) %>%
  matrix_to_df("CS_cell_type") %>%
  gather(key = "sample", value = "enrichment", CD4_CD8_0.005_rep_1:CD4_CD8_0_rep_9) %>%
  left_join(sample_df)

cs_pval_df <- cs_ssgsea_df %>%
  filter(CS_cell_type == "T.cells.CD8") %>%
  dplyr::select(sample_n, enrichment, CD8_fraction) %>%
  mutate(CD8_fraction = str_c("CD8_fraction_", CD8_fraction)) %>%
  spread(key = "CD8_fraction", value = enrichment) %>%
  dplyr::select(-sample_n)

cs_pvals <- cs_pval_df %>%

```

```

colnames %>%
.[-1] %>%
map(function(col) wilcox.test(cs_pval_df$CD8_fraction_0, extract2(cs_pval_df , col))) %>%
map_dbl(extract2, "p.value") %>%
set_names(colnames(cs_pval_df[-1]))

pval_df <- data_frame(
  "CD8_Fraction" = c(0.005, 0.01, 0.02, 0.05),
  "MCPCounter_Cytotoxic_lymphocytes" = mcp_CL_pvals,
  "MCPCounter_CD8_T_cells" = mcp_CD8_pvals,
  "Cibersort_T_cells_CD8" = cs_pvals)

kable(pval_df)

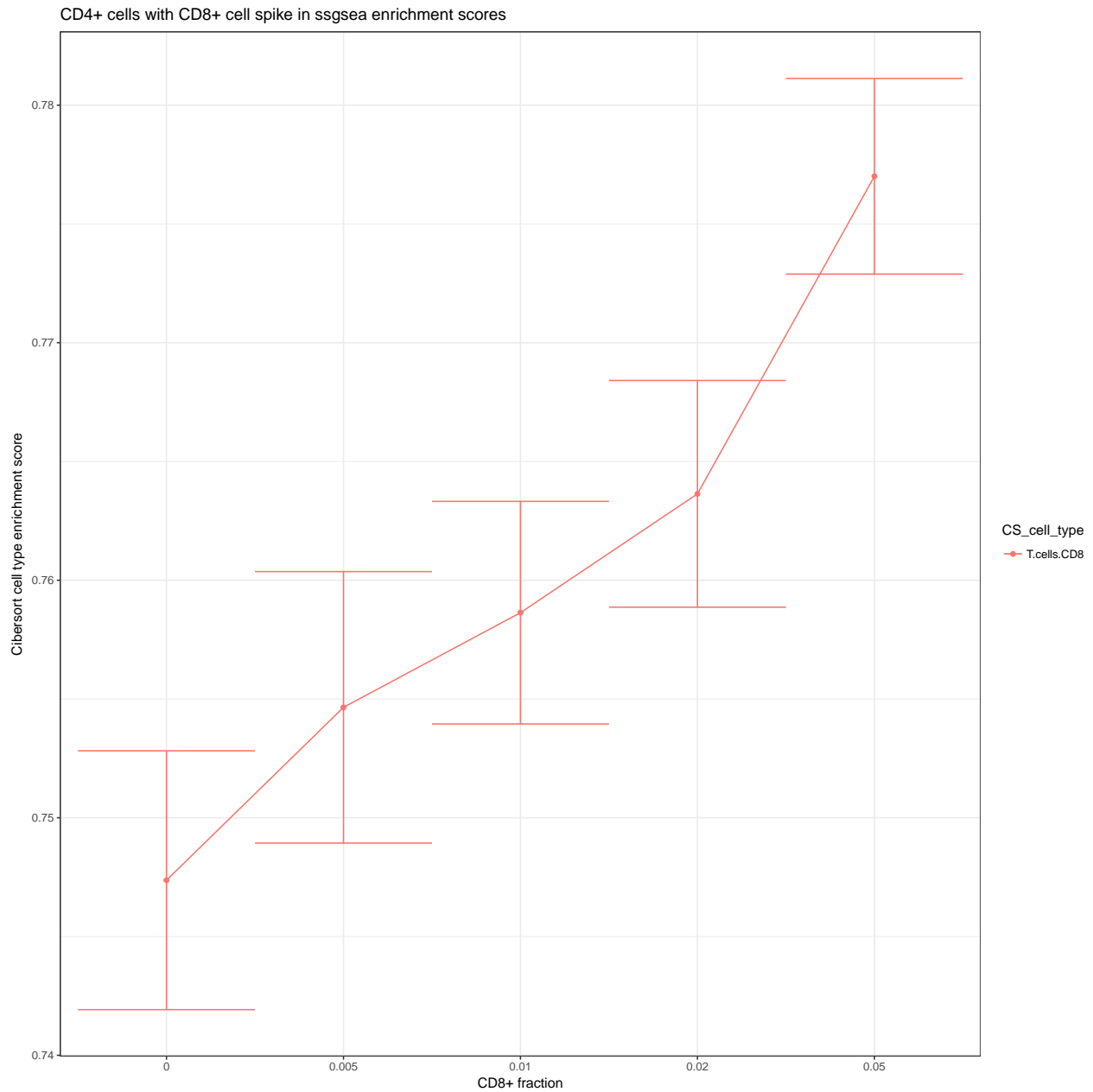
```

CD8_Fraction	MCPCounter_Cytotoxic_lymphocytes	MCPCounter_CD8_T_cells	Cibersort_T_cells_CD8
0.005	0.0015047	0.1051224	0.0185434
0.010	0.0007253	0.0015047	0.0003248
0.020	0.0000433	0.0002057	0.0000108
0.050	0.0000108	0.0000108	0.0000108

```

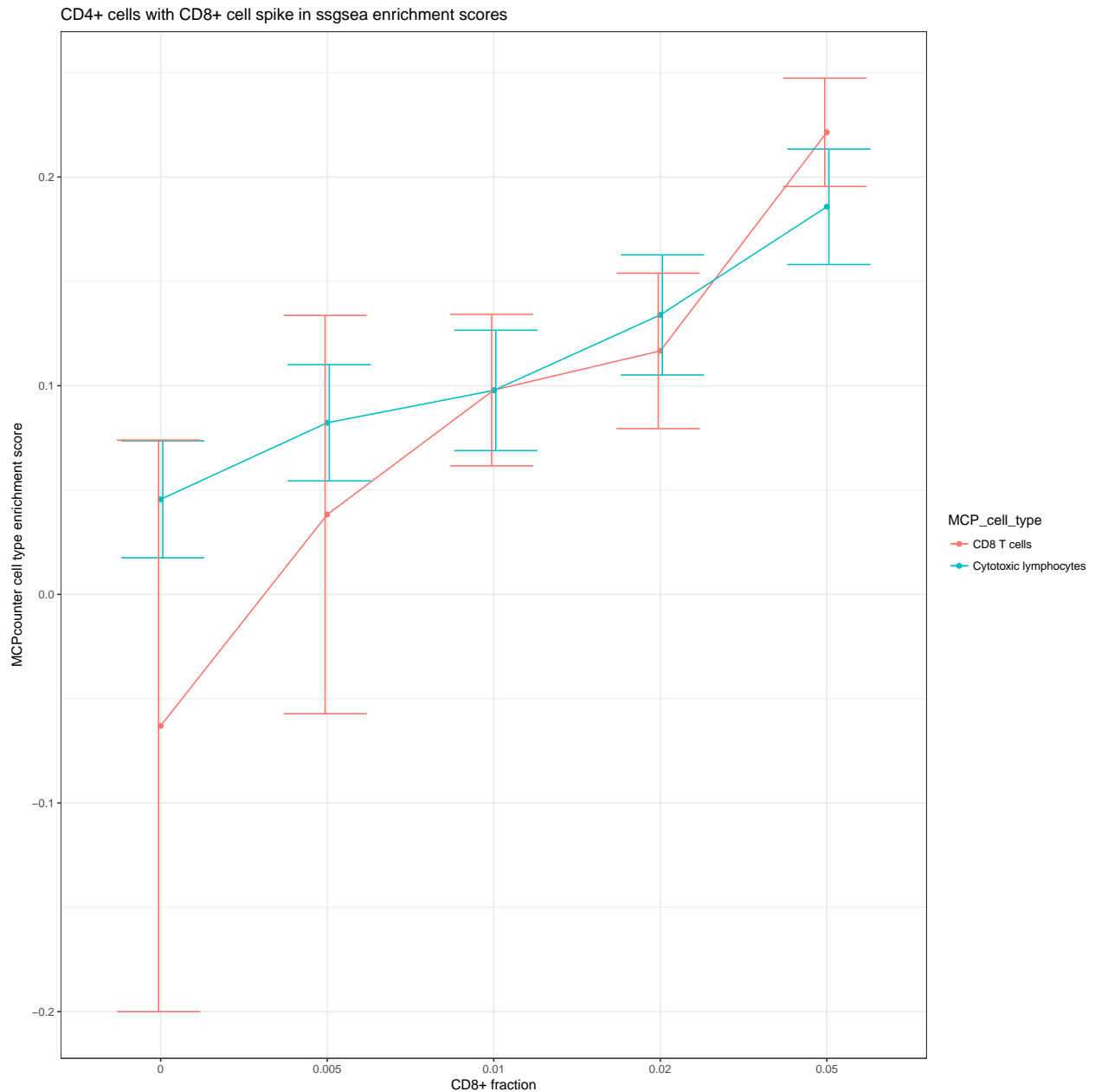
cs_ssgsea_df %>%
  filter(CS_cell_type == "T.cells.CD8") %>%
  group_by(CS_cell_type, CD8_fraction) %>%
  summarise(stdev = sd(enrichment), mean = mean(enrichment)) %>%
  ggplot(aes(x = CD8_fraction, y = mean, group = as.factor(CS_cell_type), color = CS_cell_type)) +
  geom_line() +
  geom_point() +
  theme_bw() +
  xlab("CD8+ fraction") +
  ylab("Cibersort cell type enrichment score") +
  ggtitle("CD4+ cells with CD8+ cell spike in ssgsea enrichment scores") +
  geom_errorbar(aes(ymin = mean - stdev,
                    ymax = mean + stdev),
                width = 1,
                position = position_dodge(0.05))

```



```
mcp_ssgsea_df %>%
  filter(MCP_cell_type %in% c("Cytotoxic lymphocytes", "CD8 T cells")) %>%
  group_by(MCP_cell_type, CD8_fraction) %>%
  summarise(stdev = sd(enrichment), mean = mean(enrichment)) %>%
  ggplot(aes(x = CD8_fraction, y = mean, group = as.factor(MCP_cell_type), color = MCP_cell_type)) +
  geom_line() +
  geom_point() +
  theme_bw() +
  xlab("CD8+ fraction") +
  ylab("MCPcounter cell type enrichment score") +
  ggtitle("CD4+ cells with CD8+ cell spike in ssgsea enrichment scores") +
  geom_errorbar(aes(ymin = mean - stdev,
                    ymax = mean + stdev),
                width = 1,
```

```
position = position_dodge(0.05))
```



```
cs_results_id <- "syn12385636"
mcp_results_id <- "syn12385637"

cs_result_df <- cs_results_id %>%
  create_df_from_synapse_id %>%
  dplyr::select(sample, `T cells CD8`) %>%
  gather("cibersort_cell_type", "predicted_fraction", `T cells CD8`) %>%
  full_join(sample_df, by = c("sample"))

cs_pval_df <- cs_result_df %>%
  dplyr::select(sample_n, predicted_fraction, CD8_fraction) %>%
  mutate(CD8_fraction = str_c("CD8_fraction_", CD8_fraction)) %>%
```



```

spread(key = "CD8_fraction", value = predicted_fraction) %>%
dplyr::select(-sample_n)

cs_pvals <- cs_pval_df %>%
  colnames %>%
  .[-1] %>%
  map(function(col) wilcox.test(cs_pval_df$CD8_fraction_0, extract2(cs_pval_df, col))) %>%
  map_dbl(extract2, "p.value") %>%
  set_names(colnames(cs_pval_df[-1]))

mcp_result_df <- mcp_results_id %>%
  download_from_synapse %>%
  read.table %>%
  t %>%
  matrix_to_df("sample") %>%
  set_colnames(str_replace_all(colnames(.), "\\.", " ")) %>%
  dplyr::select(sample, `CD8 T cells`, `Cytotoxic lymphocytes`) %>%
  gather("mcpcounter_cell_type", "predicted_score", `CD8 T cells`:`Cytotoxic lymphocytes`) %>%
  full_join(sample_df, by = c("sample"))

mcp_CD8_pval_df <- mcp_result_df %>%
  filter(mcpcounter_cell_type == "CD8 T cells") %>%
  dplyr::select(sample_n, predicted_score, CD8_fraction) %>%
  mutate(CD8_fraction = str_c("CD8_fraction_", CD8_fraction)) %>%
  spread(key = "CD8_fraction", value = predicted_score) %>%
  dplyr::select(-sample_n)

mcp_CD8_pvals <- mcp_CD8_pval_df %>%
  colnames %>%
  .[-1] %>%
  map(function(col) wilcox.test(mcp_CD8_pval_df$CD8_fraction_0, extract2(mcp_CD8_pval_df , col))) %>%
  map_dbl(extract2, "p.value") %>%
  set_names(colnames(mcp_CD8_pval_df[-1]))

mcp_CL_pval_df <- mcp_result_df %>%
  filter(mcpcounter_cell_type == "Cytotoxic lymphocytes") %>%
  dplyr::select(sample_n, predicted_score, CD8_fraction) %>%
  mutate(CD8_fraction = str_c("CD8_fraction_", CD8_fraction)) %>%
  spread(key = "CD8_fraction", value = predicted_score) %>%
  dplyr::select(-sample_n)

mcp_CL_pvals <- mcp_CL_pval_df %>%
  colnames %>%
  .[-1] %>%
  map(function(col) wilcox.test(mcp_CL_pval_df$CD8_fraction_0, extract2(mcp_CL_pval_df , col))) %>%
  map_dbl(extract2, "p.value") %>%
  set_names(colnames(mcp_CL_pval_df[-1]))

pval_df <- data_frame(
  "CD8_Fraction" = c(0.005, 0.01, 0.02, 0.05),
  "MCPCounter_Cytotoxic_lymphocytes" = mcp_CL_pvals,
  "MCPCounter_CD8_T_cells" = mcp_CD8_pvals,

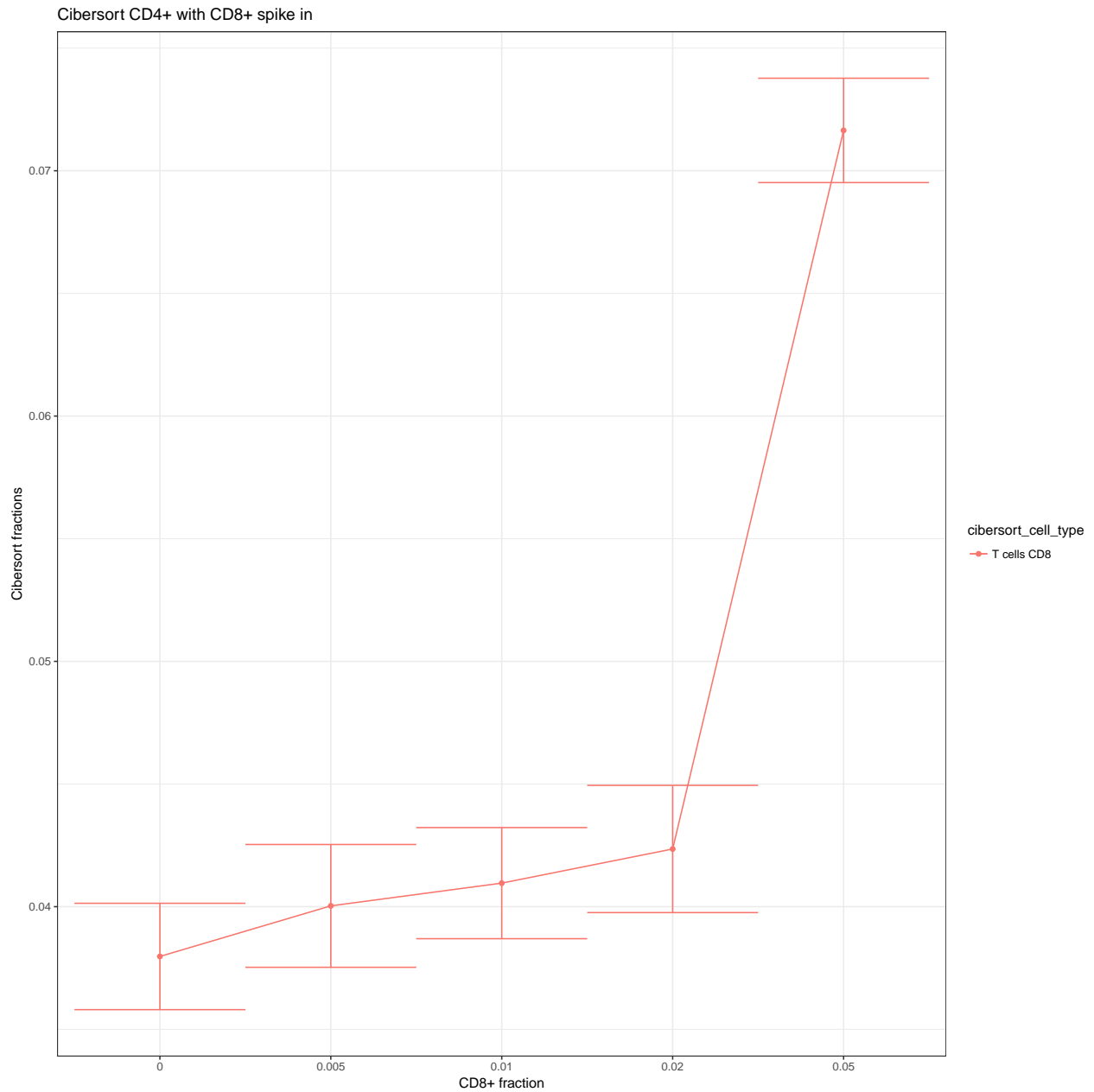
```

```
"Cibersort_T_cells_CD8" = cs_pvals)
```

```
kable(pval_df)
```

CD8_Fraction	MCPCounter_Cytotoxic_lymphocytes	MCPCounter_CD8_T_cells	Cibersort_T_cells_CD8
0.005	1.08e-05	0.1039797	0.0630128
0.010	1.08e-05	0.0028169	0.0146896
0.020	1.08e-05	0.0007650	0.0010500
0.050	1.08e-05	0.0001817	0.0000108

```
cs_result_df %>%
  group_by(cibersort_cell_type, CD8_fraction) %>%
  summarise(stdev = sd(predicted_fraction), mean = mean(predicted_fraction)) %>%
  ggplot(aes(x = CD8_fraction, y = mean, color = cibersort_cell_type)) +
  geom_line() +
  geom_point() +
  theme_bw() +
  xlab("CD8+ fraction") +
  ylab("Cibersort fractions") +
  ggtitle("Cibersort CD4+ with CD8+ spike in") +
  geom_errorbar(aes(ymin = mean - stdev,
                    ymax = mean + stdev),
                width = 1,
                position = position_dodge(0.05))
```



```
mcp_result_df %>%
  group_by(mcpcounter_cell_type, CD8_fraction) %>%
  summarise(stdev = sd(predicted_score), mean = mean(predicted_score)) %>%
  ggplot(aes(x = CD8_fraction, y = mean, color = mcpcounter_cell_type)) +
  geom_line() +
  geom_point() +
  theme_bw() +
  xlab("CD8+ fraction") +
  ylab("MCPcounter scores") +
  ggtitle("MCPCounter CD4+ with CD8+ spike in") +
  geom_errorbar(aes(ymin = mean - stdev,
                    ymax = mean + stdev),
                width = 1,
                position = position_dodge(0.05))
```

MCPCounter CD4+ with CD8+ spike in

