

Untitled

```
tpm_dir_id    <- "syn12678224"
cibersort_id  <- "syn12853640"
unsampled_reads <- "118448724"
ground_truth_id <- "syn12650217"

truth_df <- ground_truth_id %>%
  create_df_from_synapse_id %>%
  filter(`donor ID` == "HD30") %>%
  set_colnames(str_replace_all(colnames(.), " ", "_")) %>%
  select(-donor_ID)

file_df <- get_file_df_from_synapse_dir_id(tpm_dir_id)

downsampled_df <- file_df %>%
  filter(str_detect(file.name, "rep")) %>%
  mutate(reads = str_match(file.name, "([0-9]+)_rep_[0-9]+")[,2]) %>%
  mutate(rep = str_match(file.name, "[0-9]+_rep_[0-9]+")[,2]) %>%
  select(-c(file.id, rep)) %>%
  rename("sample" = file.name)

reads_df <- file_df %>%
  filter(file.name == "abundance.tsv") %>%
  select(file.name) %>%
  inset("reads", value = unsampled_reads) %>%
  rename("sample" = file.name) %>%
  bind_rows(downsampled_df)

# mcp_result_df <- mcpcounter_id %>%
#   download_from_synapse %>%
#   read.table %>%
#   t %>%
#   matrix_to_df("sample") %>%
#   mutate(sample = str_remove_all(sample, "X"))

cs_result_df <- cibersort_id %>%
  create_df_from_synapse_id

cs_summary_df <- cs_result_df %>%
  left_join(reads_df) %>%
  select(-c(sample, `P-value`, Correlation, RMSE)) %>%
  mutate(B_cells = `B cells naive` + `B cells memory`) %>%
  mutate(T_cells =
    `T cells CD8` +
    `T cells CD4 naive` +
    `T cells CD4 memory resting` +
    `T cells CD4 memory activated` +
    `T cells follicular helper` +
    `T cells regulatory (Tregs)` +
    `T cells gamma delta`) %>%
```

```

mutate(NK_cells = `NK cells resting` + `NK cells activated`) %>%
mutate(Other_cells =
  `Plasma cells` +
  `Macrophages M0` +
  `Macrophages M1` +
  `Macrophages M2` +
  `Dendritic cells resting` +
  `Dendritic cells activated` +
  `Mast cells resting` +
  `Mast cells activated` +
  `Eosinophils` +
  `Neutrophils`)

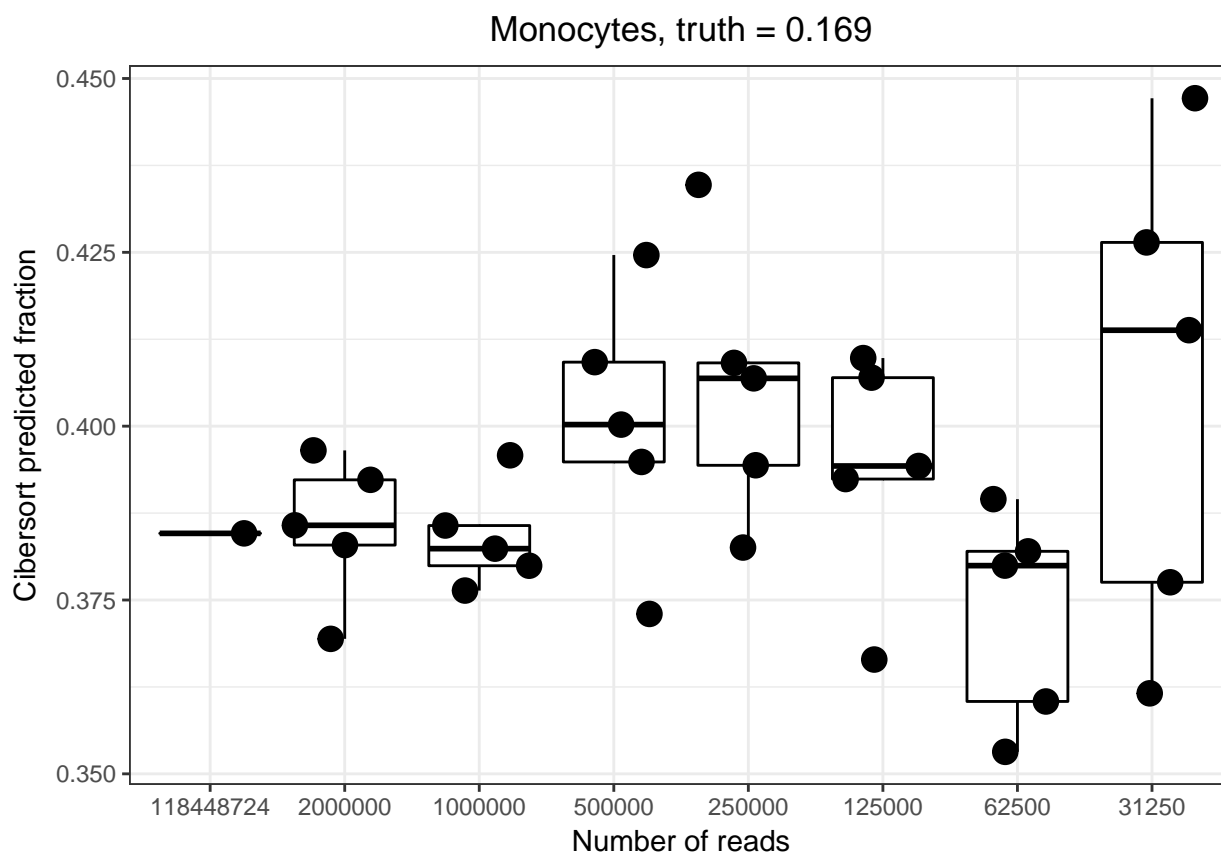
## Joining, by = "sample"
read_levels <- cs_summary_df %>%
  use_series(reads) %>%
  unique %>%
  as.integer %>%
  sort(decreasing = T) %>%
  as.character

cs_summary_df <- cs_summary_df %>%
  mutate(reads = factor(reads, levels = read_levels))

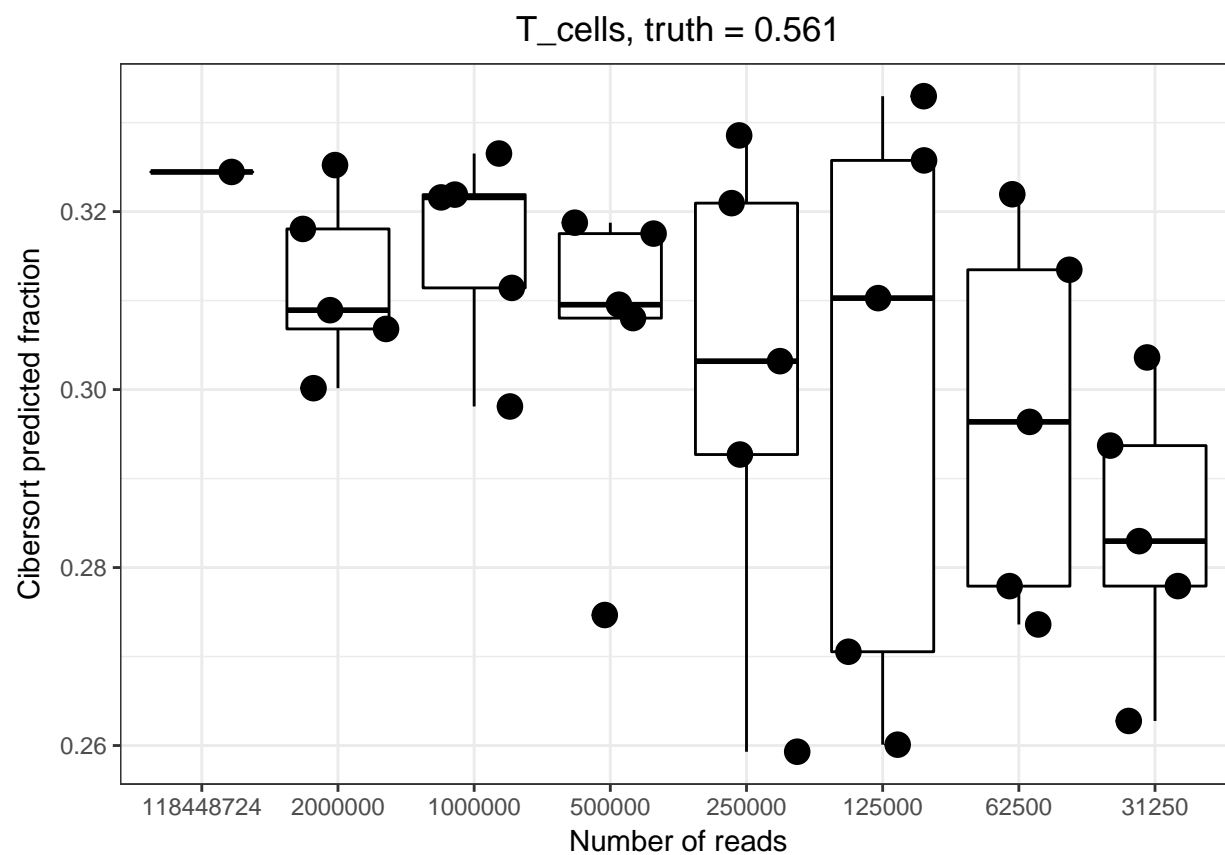
make_boxplot <- function(df, col, truth_df){
  title <- truth_df %>%
    extract2(col) %>%
    divide_by(100) %>%
    as.character %>%
    str_c(col, ", truth = ", .)
  p <- df %>%
    ggplot(aes_string(x = "reads", y = col)) +
    geom_boxplot(color = "black", fill = "white", outlier.shape = NA) +
    geom_jitter(size = 4) +
    labs(shape = "") +
    labs(color = "") +
    ggtitle(title) +
    ylab("Cibersort predicted fraction") +
    xlab("Number of reads") +
    theme_bw() +
    theme(plot.title = element_text(hjust = 0.5))
  print(p)
}

make_boxplot(cs_summary_df, "Monocytes", truth_df)

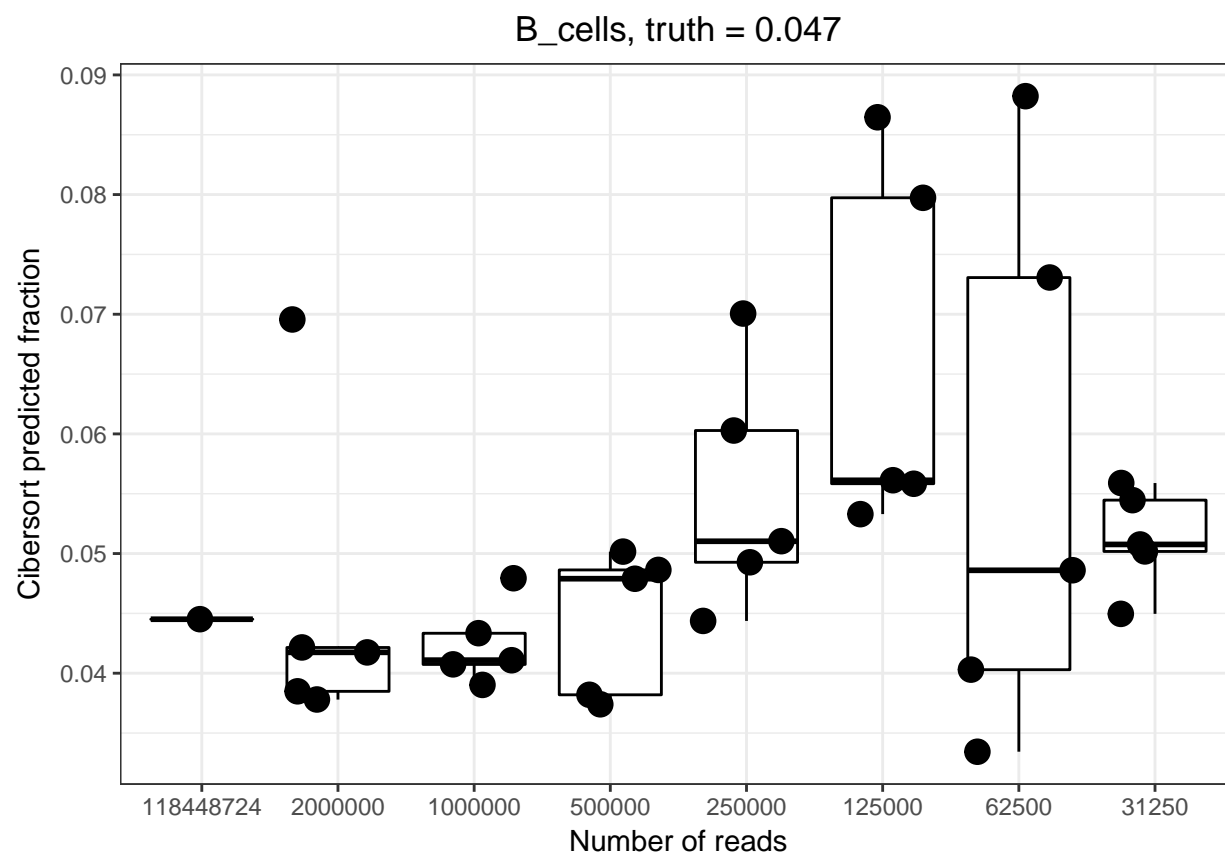
```



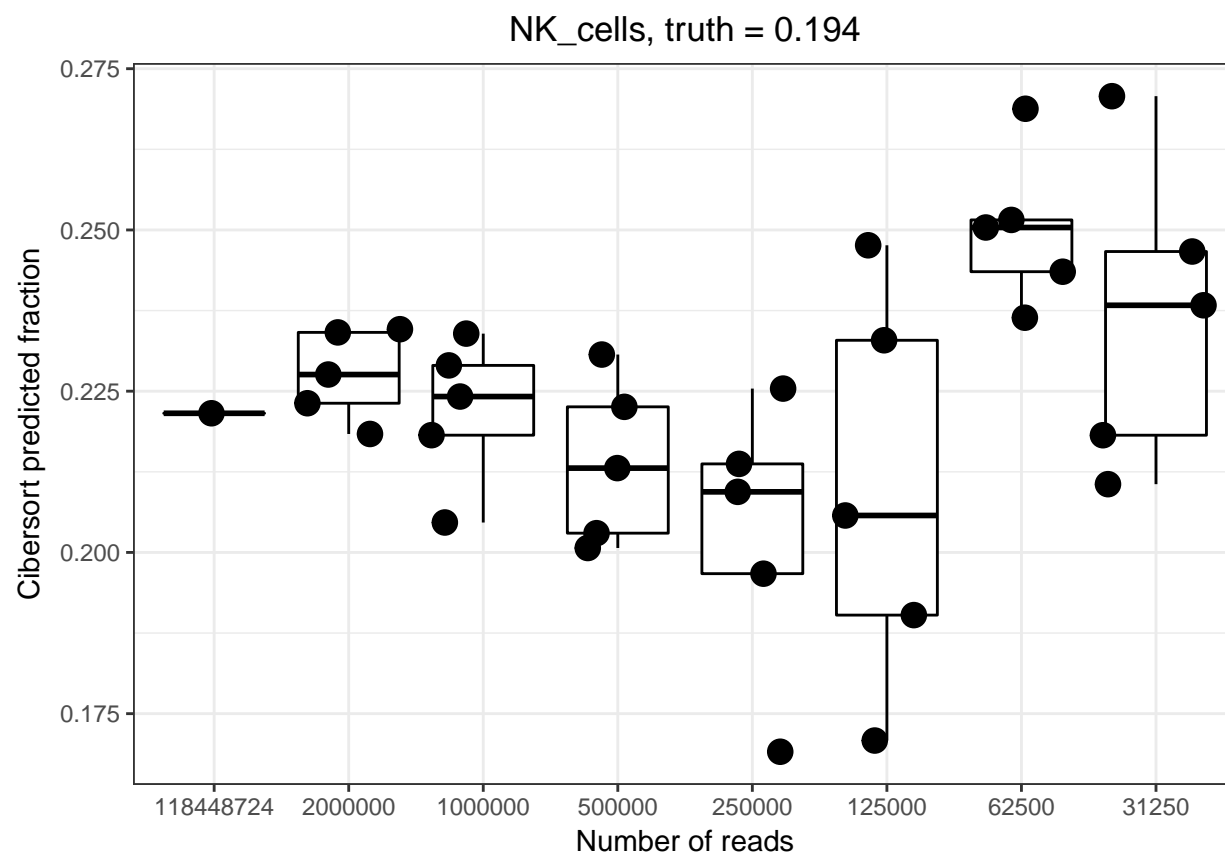
```
make_boxplot(cs_summary_df, "T_cells", truth_df)
```



```
make_boxplot(cs_summary_df, "B_cells", truth_df)
```



```
make_boxplot(cs_summary_df, "NK_cells", truth_df)
```



```
make_boxplot(cs_summary_df, "Other_cells", truth_df)
```

