

# review-expression-distributions-in-different-comparison cohorts

2023.11.16 13.58.16

hbeale

November 16, 2023

## Contents

COMPARE DISTRIBUTIONS FOR FOR OUTLIERS ACROSS COHORTS 1

expression in samples not in the compendium 3

```
outliers <- read_tsv("../input_data/druggable_outliers_from_treehouse_and_other_cohorts_2023_11_09-13_4")
mutate(high_level_cohort = ifelse(str_detect(comparison_cohort, "Treehouse"),
                                   "Treehouse",
                                   comparison_cohort))
```

```
## Rows: 287 Columns: 5
## -- Column specification -----
## Delimiter: "\t"
## chr (4): Sample_ID, comparison_cohort, gene, donor_ID
## lgl (1): pathway_support
##
## i Use `spec()` to retrieve the full column specification for this data.
## i Specify the column types or set `show_col_types = FALSE` to quiet this message.
```

## COMPARE DISTRIBUTIONS FOR FOR OUTLIERS ACROSS COHORTS

```
outlier_genes_detected <- unique(outliers$gene)

expr <- read_tsv("../input_data/druggable_TumorCompendium_v11_PolyA_hugo_log2tpm_58581genes_2020-04-09.")
rename(Sample_ID = TH_id) %>%
filter(Gene %in% outlier_genes_detected)
```

```
## Rows: 1414917 Columns: 3
## -- Column specification -----
## Delimiter: "\t"
## chr (2): Gene, TH_id
## dbl (1): log2TPM1
##
## i Use `spec()` to retrieve the full column specification for this data.
## i Specify the column types or set `show_col_types = FALSE` to quiet this message.
```

```

stanford_samples <- read_tsv("../gather_input_data/comparison_to_non_CARE_cohorts/data/TH03_TH34_rollup.sample_1.tsv")
                        col_names = "Sample_ID") %>%
mutate(cohort = "TH03_TH34")

## Rows: 110 Columns: 1
## -- Column specification -----
## Delimiter: "\t"
## chr (1): Sample_ID
##
## i Use `spec()` to retrieve the full column specification for this data.
## i Specify the column types or set `show_col_types = FALSE` to quiet this message.
TCGA_samples <- read_tsv("../gather_input_data/comparison_to_non_CARE_cohorts/data/TCGA_rollup.sample_1.tsv")
                        col_names = "Sample_ID") %>%
mutate(cohort = "TCGA")

## Rows: 9806 Columns: 1
## -- Column specification -----
## Delimiter: "\t"
## chr (1): Sample_ID
##
## i Use `spec()` to retrieve the full column specification for this data.
## i Specify the column types or set `show_col_types = FALSE` to quiet this message.
PEDAYA_samples <- read_tsv("../gather_input_data/comparison_to_non_CARE_cohorts/data/PEDAYA_rollup.sample_1.tsv")
                        col_names = "Sample_ID") %>%
mutate(cohort = "PEDAYA")

## Rows: 2814 Columns: 1
## -- Column specification -----
## Delimiter: "\t"
## chr (1): Sample_ID
##
## i Use `spec()` to retrieve the full column specification for this data.
## i Specify the column types or set `show_col_types = FALSE` to quiet this message.
pan_cancer_samples <- expr %>%
  select(Sample_ID) %>%
  distinct() %>%
  mutate(cohort = "Treehouse_pc")

samples_in_cohorts <- bind_rows(
  stanford_samples,
  TCGA_samples,
  PEDAYA_samples,
  pan_cancer_samples)

tabyl(samples_in_cohorts,
      cohort)

##      cohort      n    percent
##      PEDAYA  2814 0.11045257
##      TCGA    9806 0.38489618
##      TH03_TH34  110 0.00431762

```

```
## Treehouse_pc 12747 0.50033363
```

## expression in samples not in the compendium

```
rsem_path <- "../input_data/non_compendium_expression"

gene_name_conversion <- read_tsv(file.path(rsem_path,
                                           "EnsGeneID_Hugo_Observed_Conversions.txt"))
```

```
## Rows: 60498 Columns: 2
## -- Column specification -----
## Delimiter: "\t"
## chr (2): HugoID, EnsGeneID
##
## i Use `spec()` to retrieve the full column specification for this data.
## i Specify the column types or set `show_col_types = FALSE` to quiet this message.
```

```
relevant_gene_name_conversion <- gene_name_conversion %>%
  filter(HugoID %in% outlier_genes_detected)

rsem_kitchen_sink_data <- tibble(file_name = list.files(
  path = rsem_path,
  pattern = "_rsem_genes.results")) %>%
  rowwise() %>%
  mutate(rsem_raw = list(read_tsv(file.path(rsem_path, file_name),
                                           show_col_types = FALSE
                                           ))) %>%

  unnest(rsem_raw) %>%
  filter(gene_id %in% relevant_gene_name_conversion$EnsGeneID) %>%
  mutate(Sample_ID = str_extract(file_name, "TH[R]?[0-9]{2}_[0-9]{4}_S[0-9]{2}")) %>%
  left_join(relevant_gene_name_conversion,
            by=c("gene_id"="EnsGeneID")) %>%
  group_by(Sample_ID, HugoID) %>%
  summarize(sum_TPM = sum(TPM),
            n=n()) %>%
  mutate(log2TPM1 = log2(sum_TPM +1))
```

```
## `summarise()` has grouped output by 'Sample_ID'. You can override using the
## `.groups` argument.
```

```
table(rsem_kitchen_sink_data$n)
```

```
##
##    1    2
## 275    5
```

```
patient_expression_from_rsem_files <- rsem_kitchen_sink_data %>%
  select(gene = HugoID,
         log2TPM1,
         Sample_ID)

patient_expression_from_compendia <- outliers %>%
  select(Sample_ID, gene) %>%
  distinct() %>%
```

```

left_join(expr,
           by=c("Sample_ID", "gene"="Gene"))

patient_expression <- bind_rows(
  patient_expression_from_rsem_files,
  patient_expression_from_compendia)

length(outlier_genes_detected)

## [1] 56

outliers$Sample_ID[ ! outliers$Sample_ID %in% expr$Sample_ID] %>% unique()

## [1] "TH34_1400_S01" "TH34_2292_S01" "TH34_2666_S01" "TH34_1445_S02"
## [5] "TH34_1456_S02"

outliers
patient_expression

## # A tibble: 410 x 3
## # Groups:   Sample_ID [34]
##   gene log2TPM1 Sample_ID
##   <chr>    <dbl> <chr>
## 1 AKT1      6.41 TH34_1400_S01
## 2 AKT2      7.55 TH34_1400_S01
## 3 ALK       0.791 TH34_1400_S01
## 4 BCL6      6.68 TH34_1400_S01
## 5 BTK       2.09 TH34_1400_S01
## 6 CCND1     5.10 TH34_1400_S01
## 7 CCND2     3.35 TH34_1400_S01
## 8 CCND3     4.52 TH34_1400_S01
## 9 CCNE1     1.17 TH34_1400_S01
## 10 CDK4     5.63 TH34_1400_S01
## # i 400 more rows

# How many colors to i need

outliers %>%
  group_by(gene) %>%
  summarize(n_samples = length(unique(Sample_ID))) %>%
  arrange(desc(n_samples))

## # A tibble: 56 x 2
##   gene n_samples
##   <chr>    <int>
## 1 IGF2      18
## 2 HMOX1      8
## 3 NTRK2      7
## 4 FGFR4      5
## 5 ETV1       4
## 6 NTRK3      4
## 7 BTK        3
## 8 CDK9        3
## 9 FGFR1        3
## 10 FLT4        3

```

```
## # i 46 more rows

lapply(outlier_genes_detected, function(this_gene){
  # this_gene <- "PTCH1"
  relevant_patient_expression <- patient_expression %>%
    filter(gene == this_gene) %>%
    filter(Sample_ID %in% (outliers %>%
                          filter(gene == this_gene) %>%
                          pull(Sample_ID)))

  one_gene_expr_per_cohort <- left_join(samples_in_cohorts,
                                       expr %>%
                                       filter(Gene == this_gene))

  outlier_table <- outliers %>%
    select(Sample_ID, gene, comparison_cohort) %>%
    mutate(found = TRUE) %>%
    pivot_wider(names_from = comparison_cohort,
                values_from = found,
                values_fill = FALSE) %>%
    filter(gene == this_gene) %>%
    select(-Treehouse_pd) %>%
    left_join(relevant_patient_expression,
              by = c("Sample_ID", "gene")) %>%
    mutate(log2TPM1 = round(log2TPM1, 3))

  outlier_table_long <- outlier_table %>%
    pivot_longer(cols = c(-Sample_ID, -log2TPM1, -gene),
                 names_to = "cohort",
                 values_to = "outlier") %>%
    mutate(outlier_status = c("non outlier", "outlier")[1+outlier])

  p1 <- ggplot(one_gene_expr_per_cohort) +
    geom_histogram(aes(x=log2TPM1)) +
    geom_vline(data = outlier_table_long,
               aes(xintercept = log2TPM1,
                   color = Sample_ID,
                   lty = outlier_status)) +
    scale_color_brewer(palette = "Set1") +
    facet_col(~cohort, scales = "free_y") +
    ggtitle(this_gene)

  p2 <- ggplot(one_gene_expr_per_cohort) +
    geom_boxplot(aes(x=log2TPM1)) +
    geom_vline(data = outlier_table_long,
               aes(xintercept = log2TPM1,
                   color = Sample_ID,
                   lty = outlier_status)) +
    scale_color_brewer(palette = "Set1") +
    facet_col(~cohort)

  t1 <- tableGrob(outlier_table, theme=ttheme_minimal(), rows=NULL) # transform into a tableGrob

```

```

plot_grid(p1, p2, t1,
          ncol = 1)
})

## Joining, by = "Sample_ID"
## `stat_bin()` using `bins = 30`. Pick better value with `binwidth`.

## Warning: Removed 8 rows containing missing values (`geom_vline()`).
## Removed 8 rows containing missing values (`geom_vline()`).

## Joining, by = "Sample_ID"
## `stat_bin()` using `bins = 30`. Pick better value with `binwidth`.

## Warning: Removed 4 rows containing missing values (`geom_vline()`).
## Warning: Removed 4 rows containing missing values (`geom_vline()`).

## Joining, by = "Sample_ID"
## `stat_bin()` using `bins = 30`. Pick better value with `binwidth`.
## Joining, by = "Sample_ID"
## `stat_bin()` using `bins = 30`. Pick better value with `binwidth`.
## Joining, by = "Sample_ID"
## `stat_bin()` using `bins = 30`. Pick better value with `binwidth`.
## Joining, by = "Sample_ID"
## `stat_bin()` using `bins = 30`. Pick better value with `binwidth`.
## Joining, by = "Sample_ID"
## `stat_bin()` using `bins = 30`. Pick better value with `binwidth`.
## Joining, by = "Sample_ID"
## `stat_bin()` using `bins = 30`. Pick better value with `binwidth`.
## Joining, by = "Sample_ID"
## `stat_bin()` using `bins = 30`. Pick better value with `binwidth`.
## Joining, by = "Sample_ID"
## `stat_bin()` using `bins = 30`. Pick better value with `binwidth`.
## Joining, by = "Sample_ID"
## `stat_bin()` using `bins = 30`. Pick better value with `binwidth`.
## Joining, by = "Sample_ID"
## `stat_bin()` using `bins = 30`. Pick better value with `binwidth`.
## Warning: Removed 8 rows containing missing values (`geom_vline()`).
## Warning: Removed 8 rows containing missing values (`geom_vline()`).

## Joining, by = "Sample_ID"
## `stat_bin()` using `bins = 30`. Pick better value with `binwidth`.
## Joining, by = "Sample_ID"
## `stat_bin()` using `bins = 30`. Pick better value with `binwidth`.

## Warning: Removed 4 rows containing missing values (`geom_vline()`).
## Warning: Removed 4 rows containing missing values (`geom_vline()`).

## Joining, by = "Sample_ID"
## `stat_bin()` using `bins = 30`. Pick better value with `binwidth`.

## Warning: Removed 4 rows containing missing values (`geom_vline()`).
## Removed 4 rows containing missing values (`geom_vline()`).

```

```

## Joining, by = "Sample_ID"
## `stat_bin()` using `bins = 30`. Pick better value with `binwidth`.

## Warning in RColorBrewer::brewer.pal(n, pal): n too large, allowed maximum for palette Set1 is 9
## Returning the palette you asked for with that many colors

## Warning in RColorBrewer::brewer.pal(n, pal): Removed 4 rows containing missing values (`geom_vline()`)
## Warning in RColorBrewer::brewer.pal(n, pal): n too large, allowed maximum for palette Set1 is 9
## Returning the palette you asked for with that many colors
## Warning: Removed 4 rows containing missing values (`geom_vline()`).

## Joining, by = "Sample_ID"
## `stat_bin()` using `bins = 30`. Pick better value with `binwidth`.

## Warning: Removed 4 rows containing missing values (`geom_vline()`).
## Removed 4 rows containing missing values (`geom_vline()`).

## Joining, by = "Sample_ID"
## `stat_bin()` using `bins = 30`. Pick better value with `binwidth`.
## Joining, by = "Sample_ID"
## `stat_bin()` using `bins = 30`. Pick better value with `binwidth`.
## Joining, by = "Sample_ID"
## `stat_bin()` using `bins = 30`. Pick better value with `binwidth`.
## Joining, by = "Sample_ID"
## `stat_bin()` using `bins = 30`. Pick better value with `binwidth`.
## Joining, by = "Sample_ID"
## `stat_bin()` using `bins = 30`. Pick better value with `binwidth`.
## Joining, by = "Sample_ID"
## `stat_bin()` using `bins = 30`. Pick better value with `binwidth`.
## Joining, by = "Sample_ID"
## `stat_bin()` using `bins = 30`. Pick better value with `binwidth`.
## Joining, by = "Sample_ID"
## `stat_bin()` using `bins = 30`. Pick better value with `binwidth`.
## Joining, by = "Sample_ID"
## `stat_bin()` using `bins = 30`. Pick better value with `binwidth`.
## Joining, by = "Sample_ID"
## `stat_bin()` using `bins = 30`. Pick better value with `binwidth`.
## Warning: Removed 8 rows containing missing values (`geom_vline()`).
## Warning: Removed 8 rows containing missing values (`geom_vline()`).

## Joining, by = "Sample_ID"
## `stat_bin()` using `bins = 30`. Pick better value with `binwidth`.
## Joining, by = "Sample_ID"
## `stat_bin()` using `bins = 30`. Pick better value with `binwidth`.
## Joining, by = "Sample_ID"
## `stat_bin()` using `bins = 30`. Pick better value with `binwidth`.
## Joining, by = "Sample_ID"
## `stat_bin()` using `bins = 30`. Pick better value with `binwidth`.
## Joining, by = "Sample_ID"
## `stat_bin()` using `bins = 30`. Pick better value with `binwidth`.
## Joining, by = "Sample_ID"
## `stat_bin()` using `bins = 30`. Pick better value with `binwidth`.
## Joining, by = "Sample_ID"
## `stat_bin()` using `bins = 30`. Pick better value with `binwidth`.
## Joining, by = "Sample_ID"
## `stat_bin()` using `bins = 30`. Pick better value with `binwidth`.

```

```

## Warning: Removed 8 rows containing missing values (`geom_vline()`).
## Removed 8 rows containing missing values (`geom_vline()`).

## Joining, by = "Sample_ID"
## `stat_bin()` using `bins = 30`. Pick better value with `binwidth`.
## Joining, by = "Sample_ID"
## `stat_bin()` using `bins = 30`. Pick better value with `binwidth`.

## Warning: Removed 4 rows containing missing values (`geom_vline()`).
## Warning: Removed 4 rows containing missing values (`geom_vline()`).

## Joining, by = "Sample_ID"
## `stat_bin()` using `bins = 30`. Pick better value with `binwidth`.
## Joining, by = "Sample_ID"
## `stat_bin()` using `bins = 30`. Pick better value with `binwidth`.

## Warning: Removed 4 rows containing missing values (`geom_vline()`).
## Removed 4 rows containing missing values (`geom_vline()`).

## Joining, by = "Sample_ID"
## `stat_bin()` using `bins = 30`. Pick better value with `binwidth`.
## Joining, by = "Sample_ID"
## `stat_bin()` using `bins = 30`. Pick better value with `binwidth`.
## Joining, by = "Sample_ID"
## `stat_bin()` using `bins = 30`. Pick better value with `binwidth`.

## Warning: Removed 4 rows containing missing values (`geom_vline()`).
## Removed 4 rows containing missing values (`geom_vline()`).

## Joining, by = "Sample_ID"
## `stat_bin()` using `bins = 30`. Pick better value with `binwidth`.
## Joining, by = "Sample_ID"
## `stat_bin()` using `bins = 30`. Pick better value with `binwidth`.

## Warning: Removed 8 rows containing missing values (`geom_vline()`).
## Warning: Removed 8 rows containing missing values (`geom_vline()`).

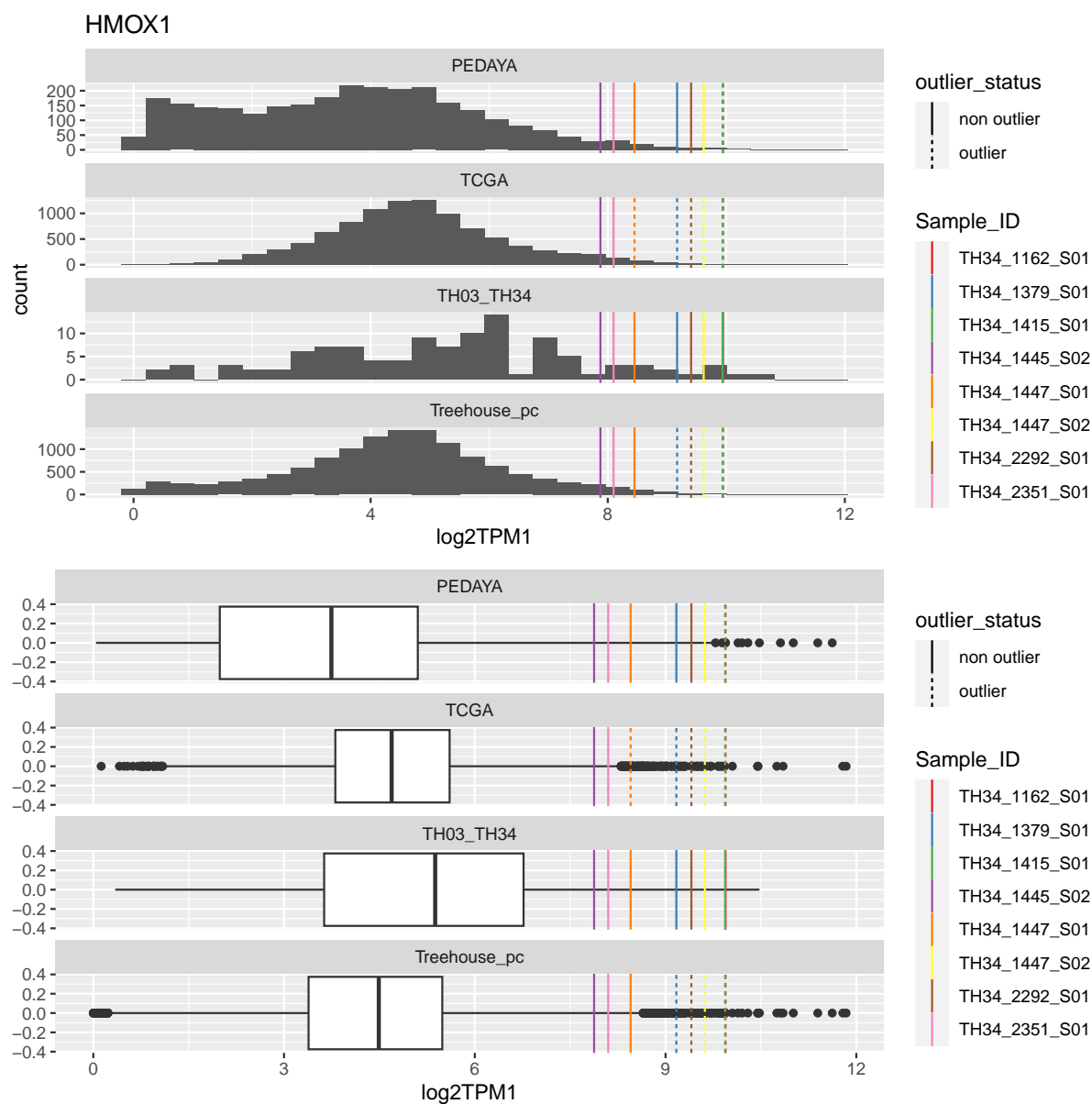
## Joining, by = "Sample_ID"
## `stat_bin()` using `bins = 30`. Pick better value with `binwidth`.
## Joining, by = "Sample_ID"
## `stat_bin()` using `bins = 30`. Pick better value with `binwidth`.
## Joining, by = "Sample_ID"
## `stat_bin()` using `bins = 30`. Pick better value with `binwidth`.
## Joining, by = "Sample_ID"
## `stat_bin()` using `bins = 30`. Pick better value with `binwidth`.
## Joining, by = "Sample_ID"
## `stat_bin()` using `bins = 30`. Pick better value with `binwidth`.
## Joining, by = "Sample_ID"
## `stat_bin()` using `bins = 30`. Pick better value with `binwidth`.
## Joining, by = "Sample_ID"
## `stat_bin()` using `bins = 30`. Pick better value with `binwidth`.
## Joining, by = "Sample_ID"
## `stat_bin()` using `bins = 30`. Pick better value with `binwidth`.

## Warning: Removed 4 rows containing missing values (`geom_vline()`).
## Warning: Removed 4 rows containing missing values (`geom_vline()`).

```

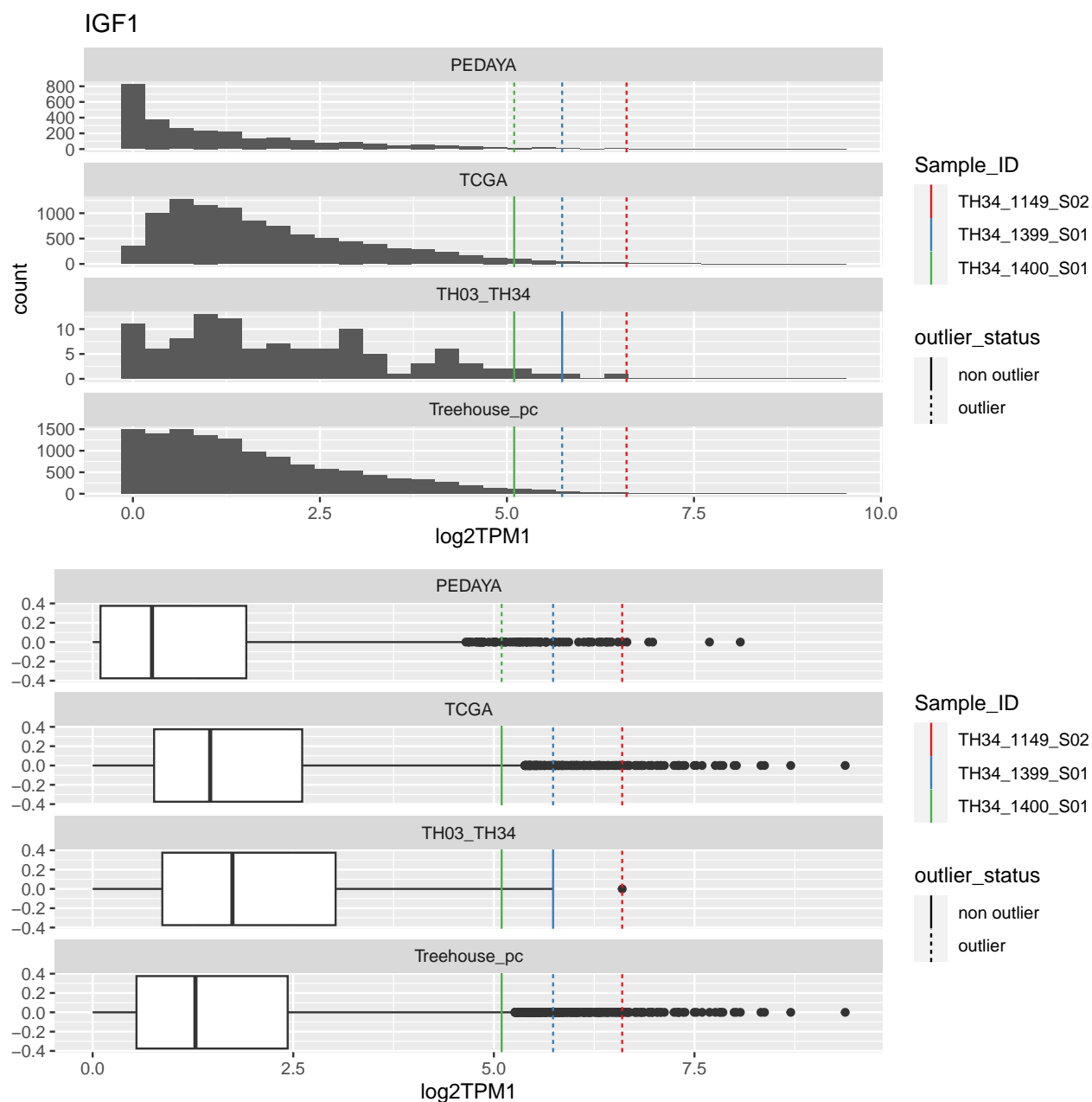


```
## Joining, by = "Sample_ID"
## `stat_bin()` using `bins = 30`. Pick better value with `binwidth`.
## Joining, by = "Sample_ID"
## `stat_bin()` using `bins = 30`. Pick better value with `binwidth`.
## Joining, by = "Sample_ID"
## `stat_bin()` using `bins = 30`. Pick better value with `binwidth`.
## Warning: Removed 4 rows containing missing values (`geom_vline()`).
## Removed 4 rows containing missing values (`geom_vline()`).
## [[1]]
```



Sample_ID	gene	PEDAYA	TCGA	TH03_TH34	Treehouse_pc	log2TPM1
TH34_1162_S01	HMOX1	TRUE	TRUE	FALSE	TRUE	9.943
TH34_1415_S01	HMOX1	TRUE	TRUE	FALSE	TRUE	9.937
TH34_1379_S01	HMOX1	FALSE	TRUE	FALSE	TRUE	9.170
TH34_1447_S01	HMOX1	FALSE	TRUE	FALSE	FALSE	8.451
TH34_1447_S02	HMOX1	FALSE	TRUE	FALSE	TRUE	9.618
TH34_2292_S01	HMOX1	FALSE	TRUE	FALSE	TRUE	9.406
TH34_2292_S01	HMOX1	FALSE	TRUE	FALSE	TRUE	NA
TH34_1445_S02	HMOX1	FALSE	FALSE	FALSE	FALSE	7.876
TH34_1445_S02	HMOX1	FALSE	FALSE	FALSE	FALSE	NA
TH34_2351_S01	HMOX1	FALSE	FALSE	FALSE	FALSE	8.098

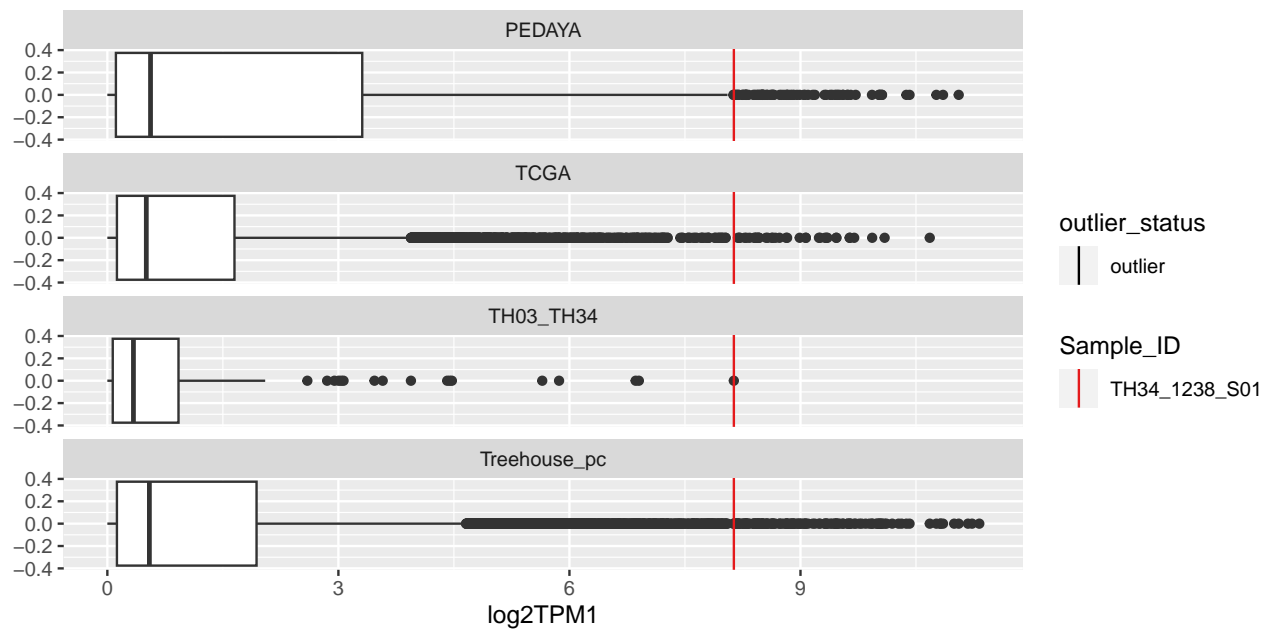
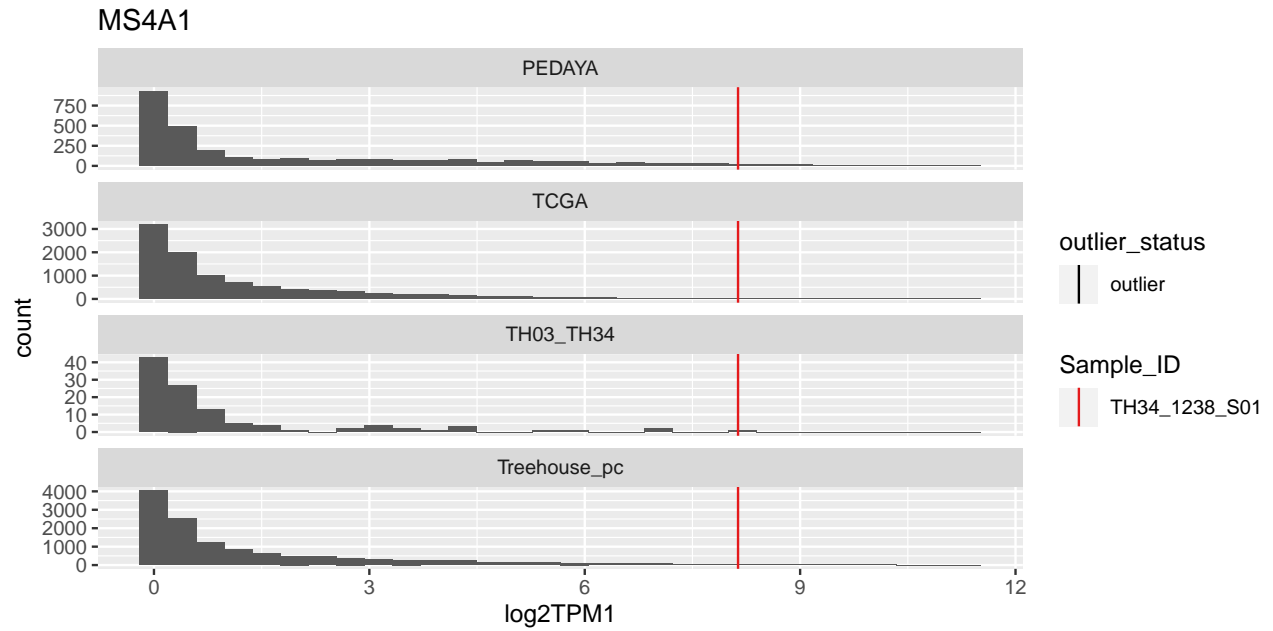
```
##  
## [[2]]
```



Sample_ID	gene	PEDAYA	TCGA	TH03_TH34	Treehouse_pc	log2TPM1
TH34_1149_S02	IGF1	TRUE	TRUE	TRUE	TRUE	6.599
TH34_1399_S01	IGF1	TRUE	TRUE	FALSE	TRUE	5.738
TH34_1400_S01	IGF1	TRUE	FALSE	FALSE	FALSE	5.097
TH34_1400_S01	IGF1	TRUE	FALSE	FALSE	FALSE	NA

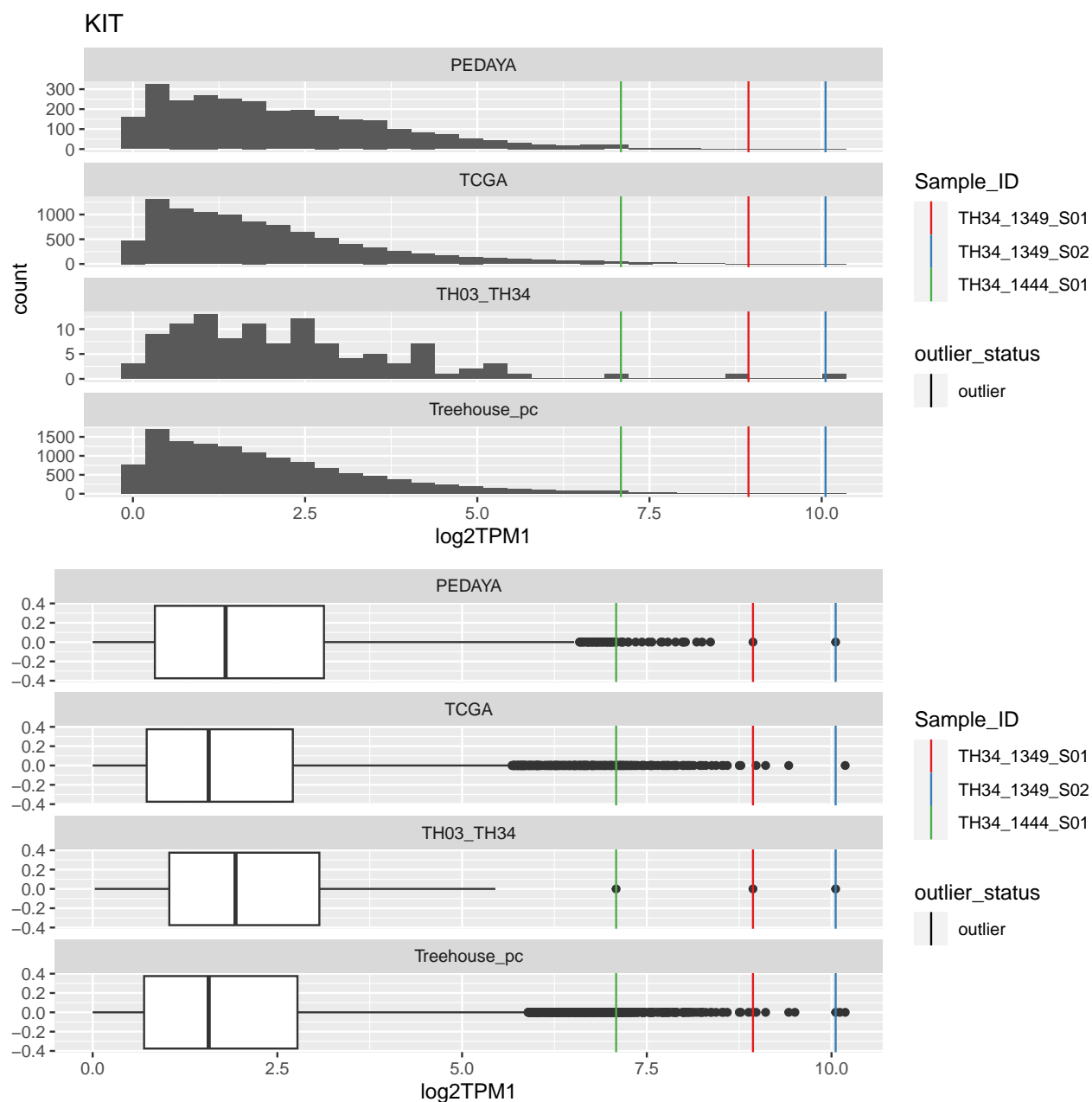
##

## [[3]]



Sample_ID	gene	PEDAYA	TCGA	TH03_TH34	Treehouse_pc	log2TPM1
TH34_1238_S01	MS4A1	TRUE	TRUE	TRUE	TRUE	8.135

```
##  
## [[4]]
```

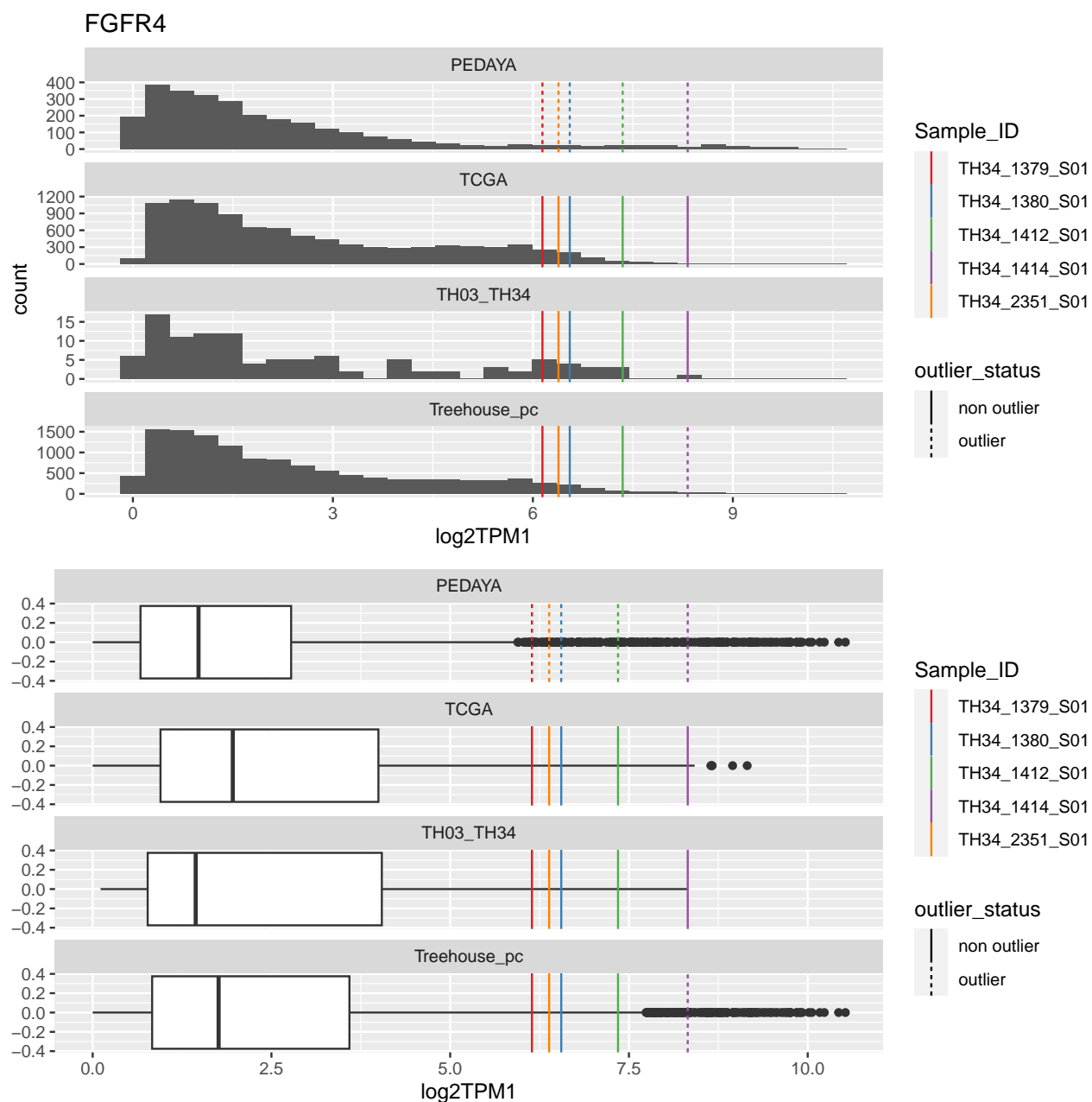


Sample_ID	gene	PEDAYA	TCGA	TH03_TH34	Treehouse_pc	log2TPM1
TH34_1349_S01	KIT	TRUE	TRUE	TRUE	TRUE	8.937
TH34_1349_S02	KIT	TRUE	TRUE	TRUE	TRUE	10.056
TH34_1444_S01	KIT	TRUE	TRUE	TRUE	TRUE	7.085

##

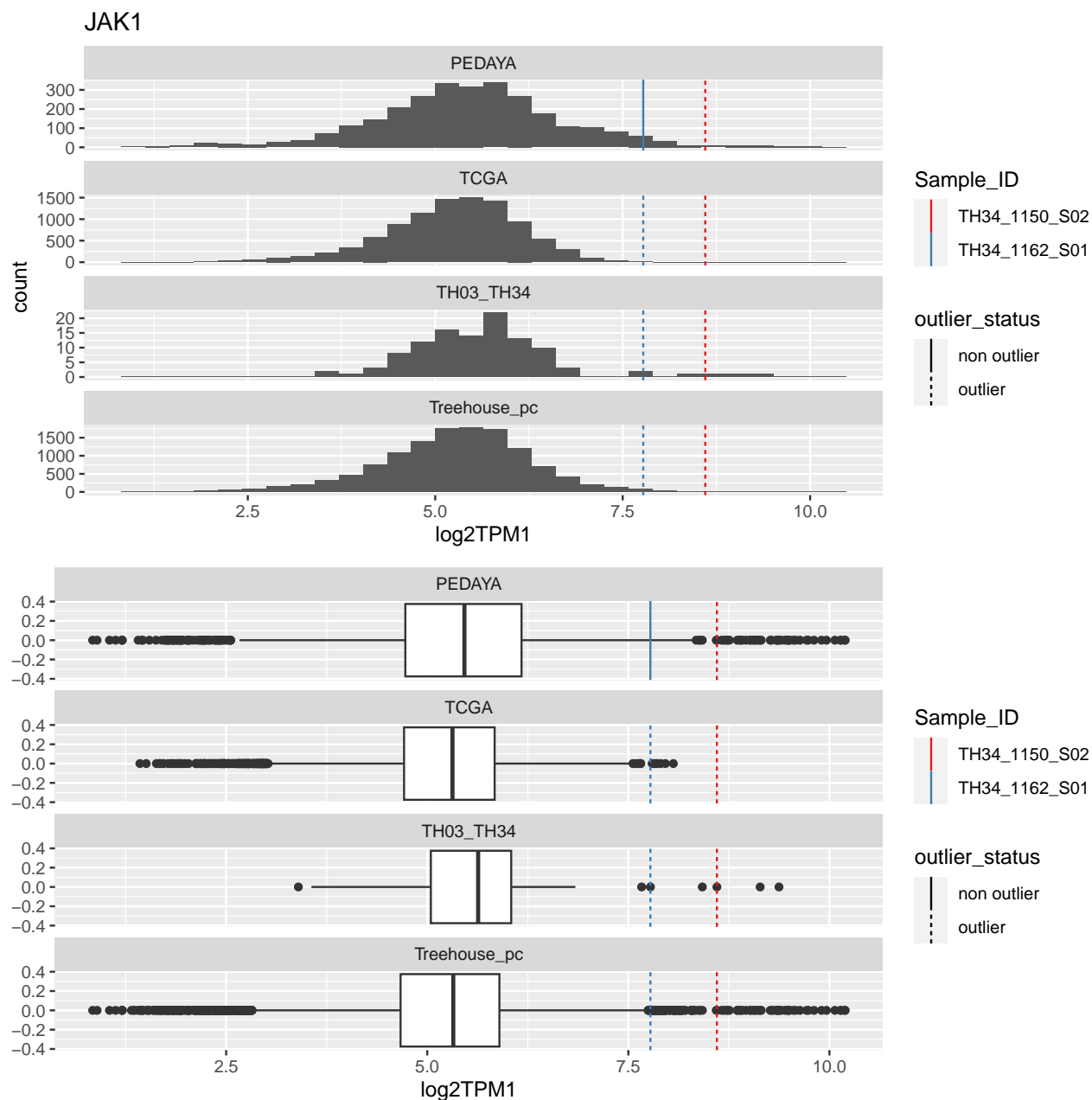
## [[5]]





Sample_ID	gene	PEDAYA	TCGA	TH03_TH34	Treehouse_pc	log2TPM1
TH34_1379_S01	FGFR4	TRUE	FALSE	FALSE	FALSE	6.143
TH34_1380_S01	FGFR4	TRUE	FALSE	FALSE	FALSE	6.553
TH34_1412_S01	FGFR4	TRUE	FALSE	FALSE	FALSE	7.347
TH34_1414_S01	FGFR4	TRUE	FALSE	FALSE	TRUE	8.322
TH34_2351_S01	FGFR4	TRUE	FALSE	FALSE	FALSE	6.384

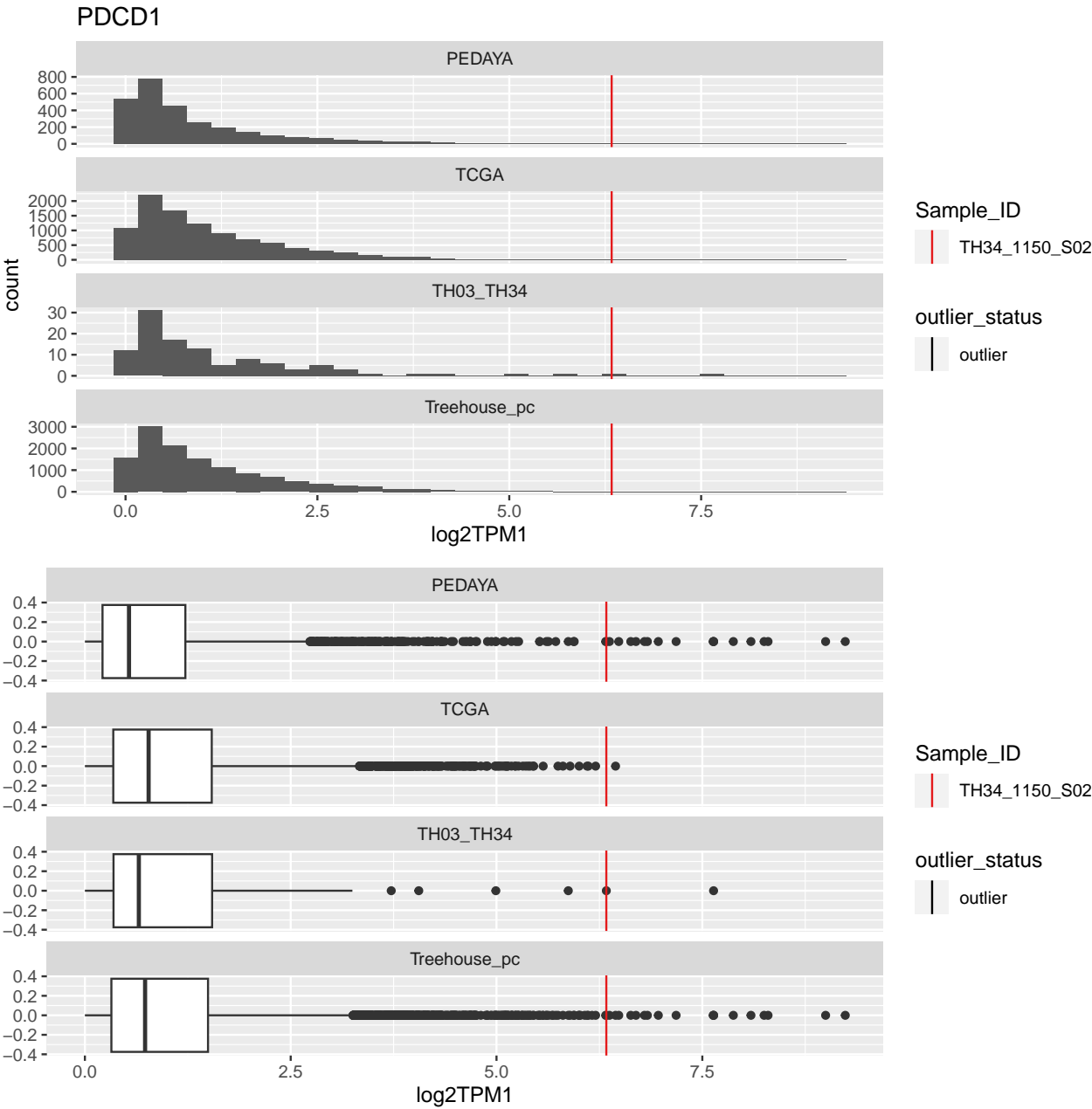
```
##  
## [[6]]
```



Sample_ID	gene	PEDAYA	TCGA	TH03_TH34	Treehouse_pc	log2TPM1
TH34_1150_S02	JAK1	TRUE	TRUE	TRUE	TRUE	8.601
TH34_1162_S01	JAK1	FALSE	TRUE	TRUE	TRUE	7.774

##

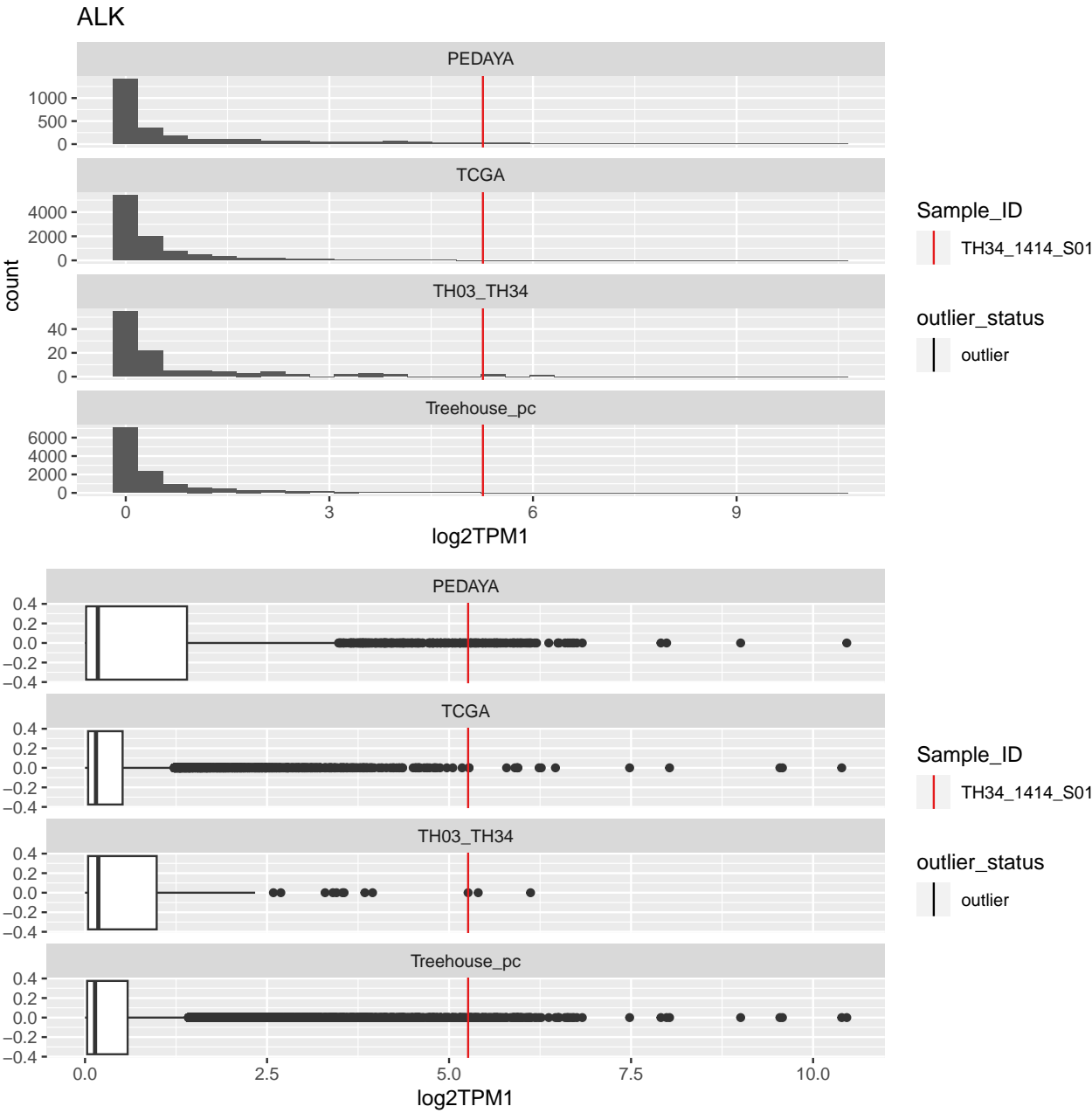
## [[7]]



Sample_ID	gene	PEDAYA	TCGA	TH03_TH34	Treehouse_pc	log2TPM1
TH34_1150_S02	PDCD1	TRUE	TRUE	TRUE	TRUE	6.334

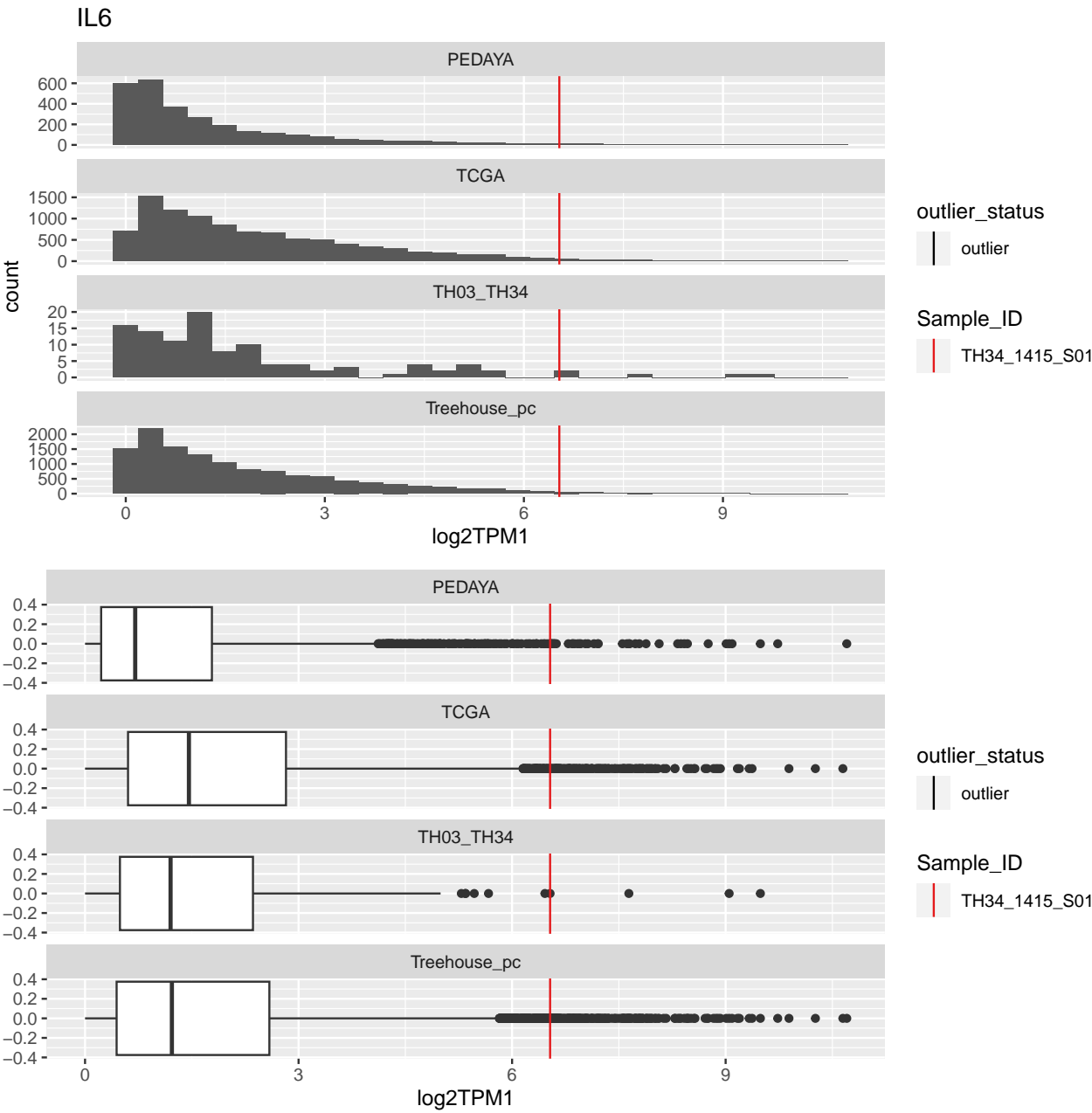
##

## [[8]]



Sample_ID	gene	PEDAYA	TCGA	TH03_TH34	Treehouse_pc	log2TPM1
TH34_1414_S01	ALK	TRUE	TRUE	TRUE	TRUE	5.262

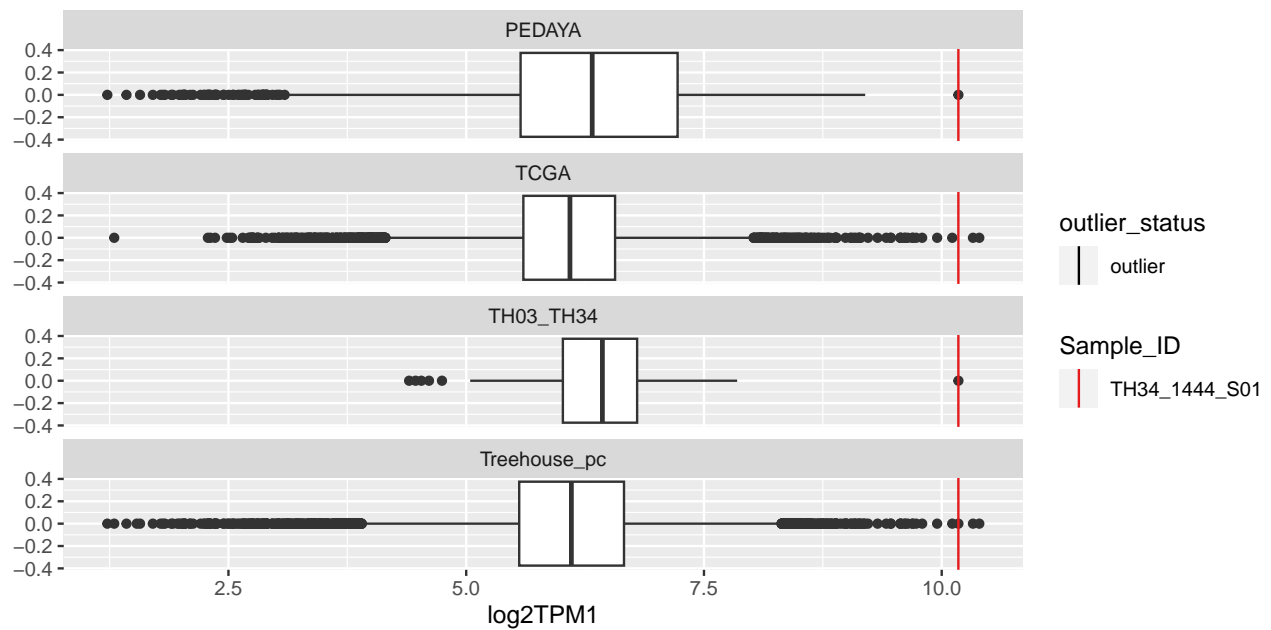
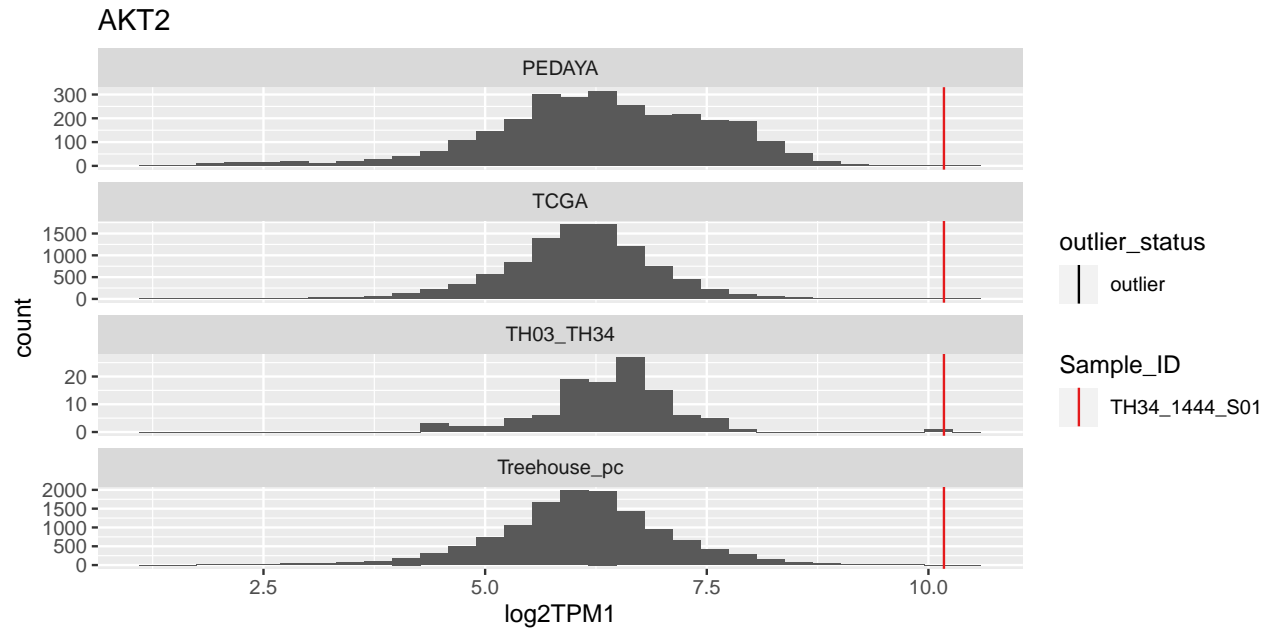
##  
## [[9]]



Sample_ID	gene	PEDAYA	TCGA	TH03_TH34	Treehouse_pc	log2TPM1
TH34_1415_S01	IL6	TRUE	TRUE	TRUE	TRUE	6.534

##

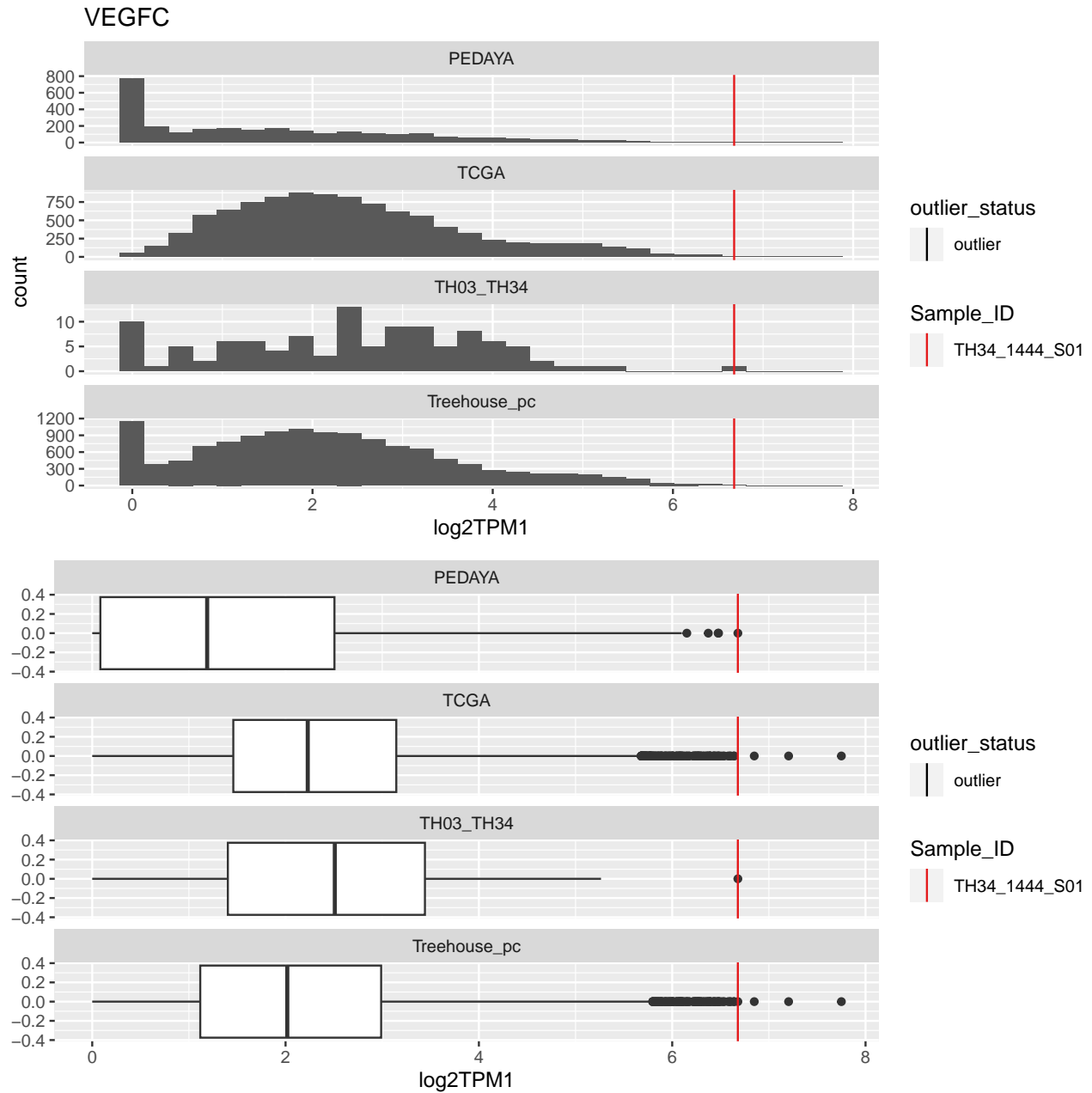
## [[10]]



Sample_ID	gene	PEDAYA	TCGA	TH03_TH34	Treehouse_pc	log2TPM1
TH34_1444_S01	AKT2	TRUE	TRUE	TRUE	TRUE	10.175

##

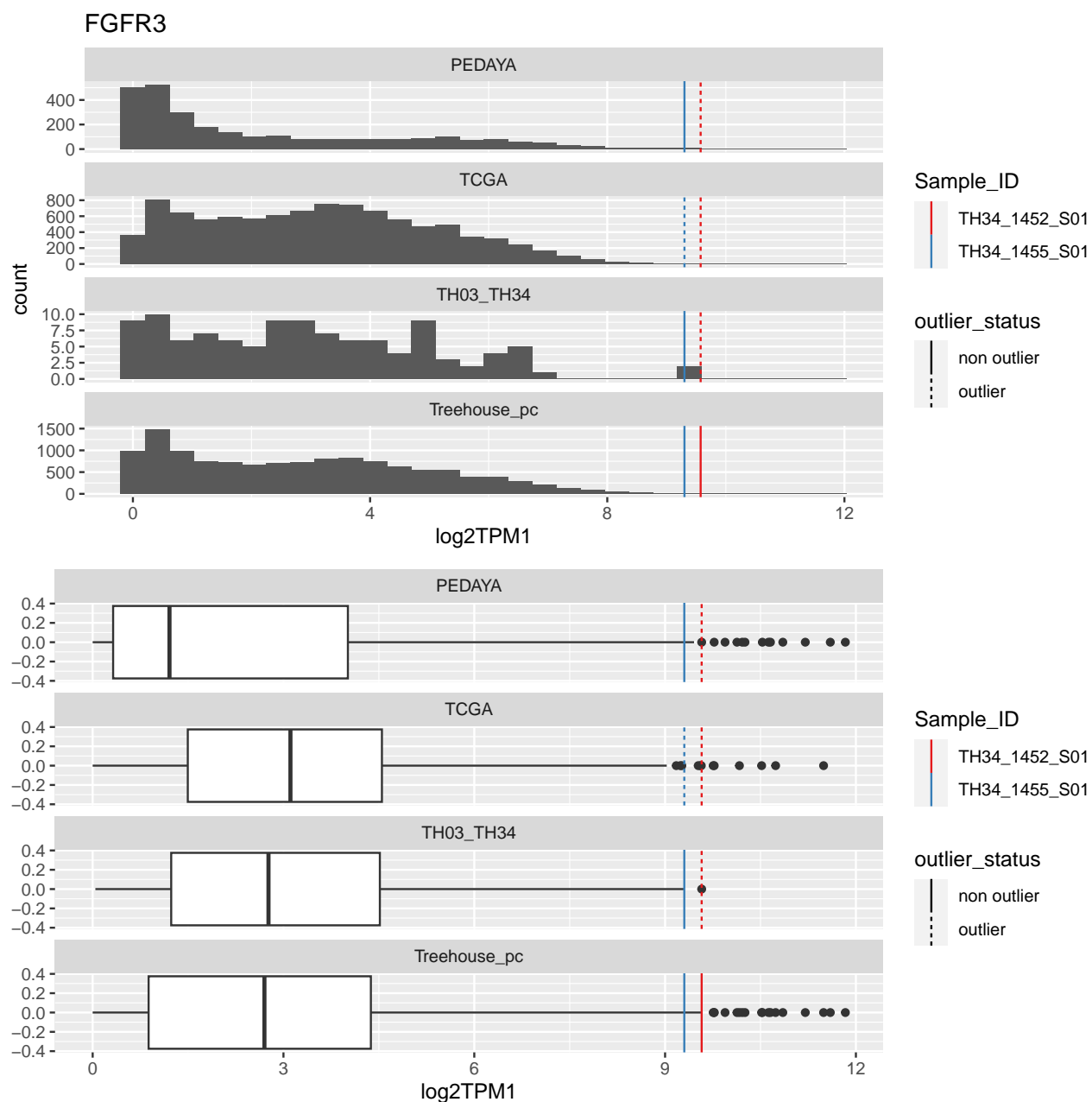
## [[11]]



Sample_ID	gene	PEDAYA	TCGA	TH03_TH34	Treehouse_pc	log2TPM1
TH34_1444_S01	VEGFC	TRUE	TRUE	TRUE	TRUE	6.679



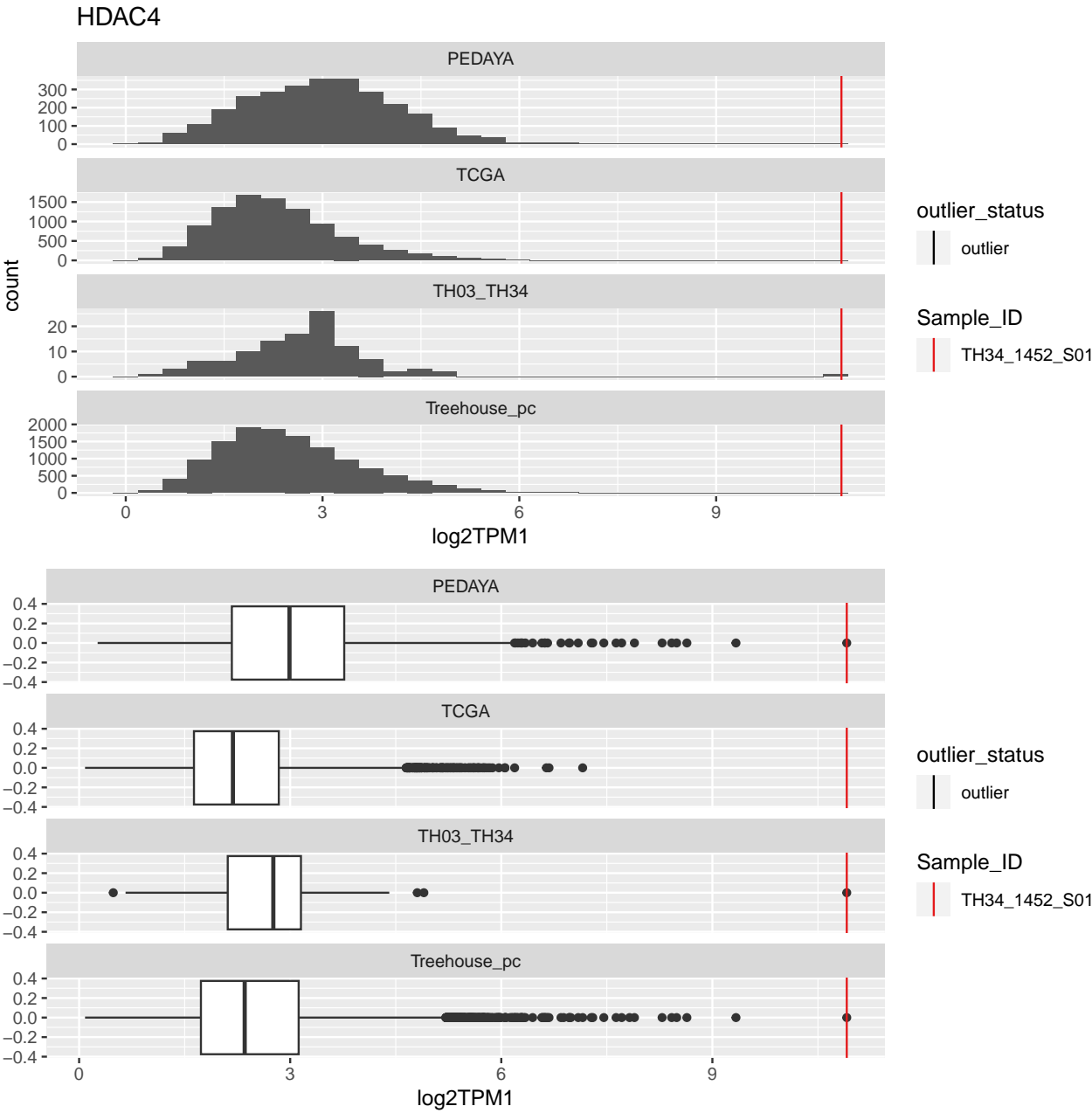
```
##  
## [[12]]
```



Sample_ID	gene	PEDAYA	TCGA	TH03_TH34	Treehouse_pc	log2TPM1
TH34_1452_S01	FGFR3	TRUE	TRUE	TRUE	FALSE	9.575
TH34_1455_S01	FGFR3	FALSE	TRUE	FALSE	FALSE	9.302

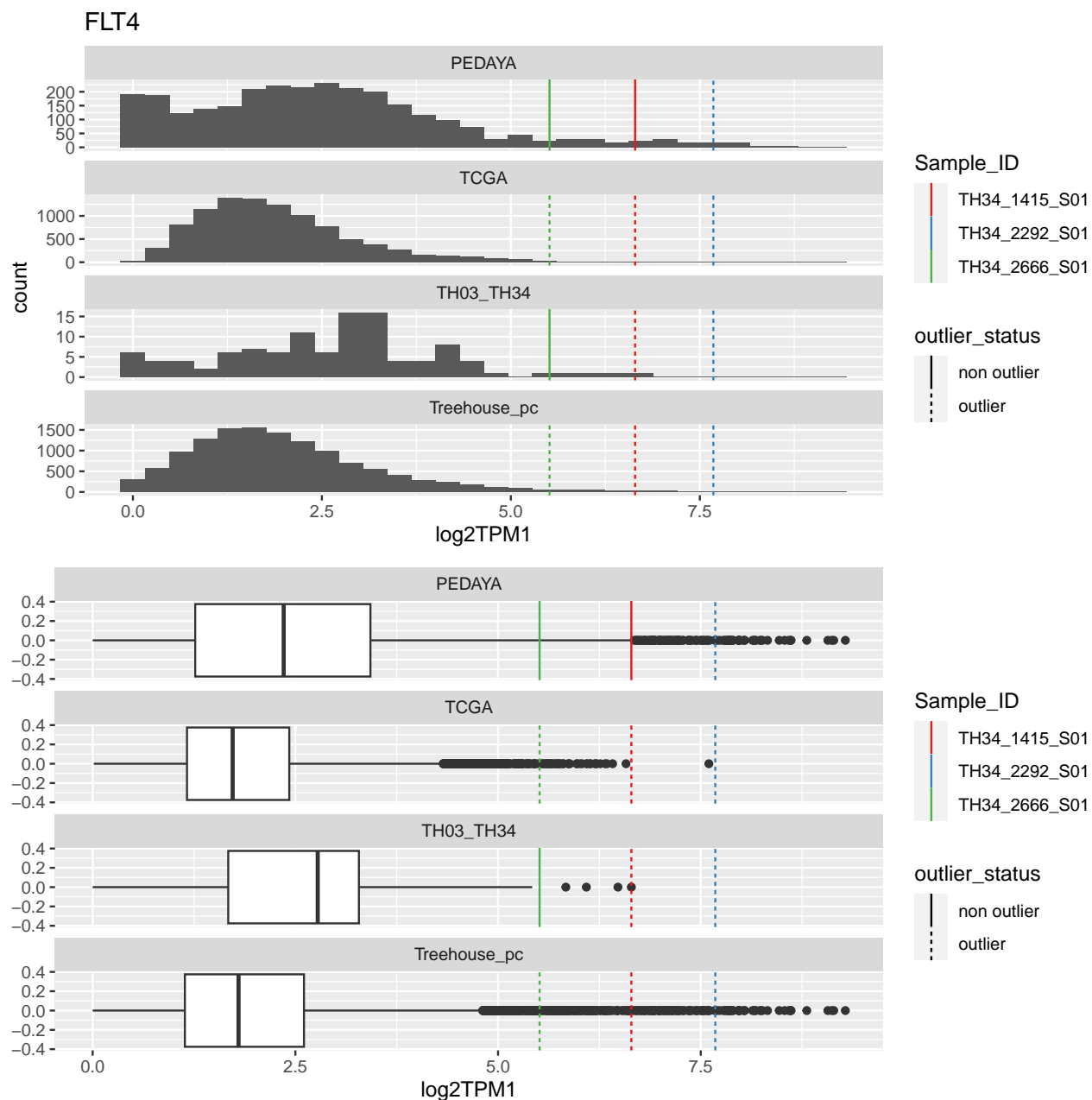
##

## [[13]]



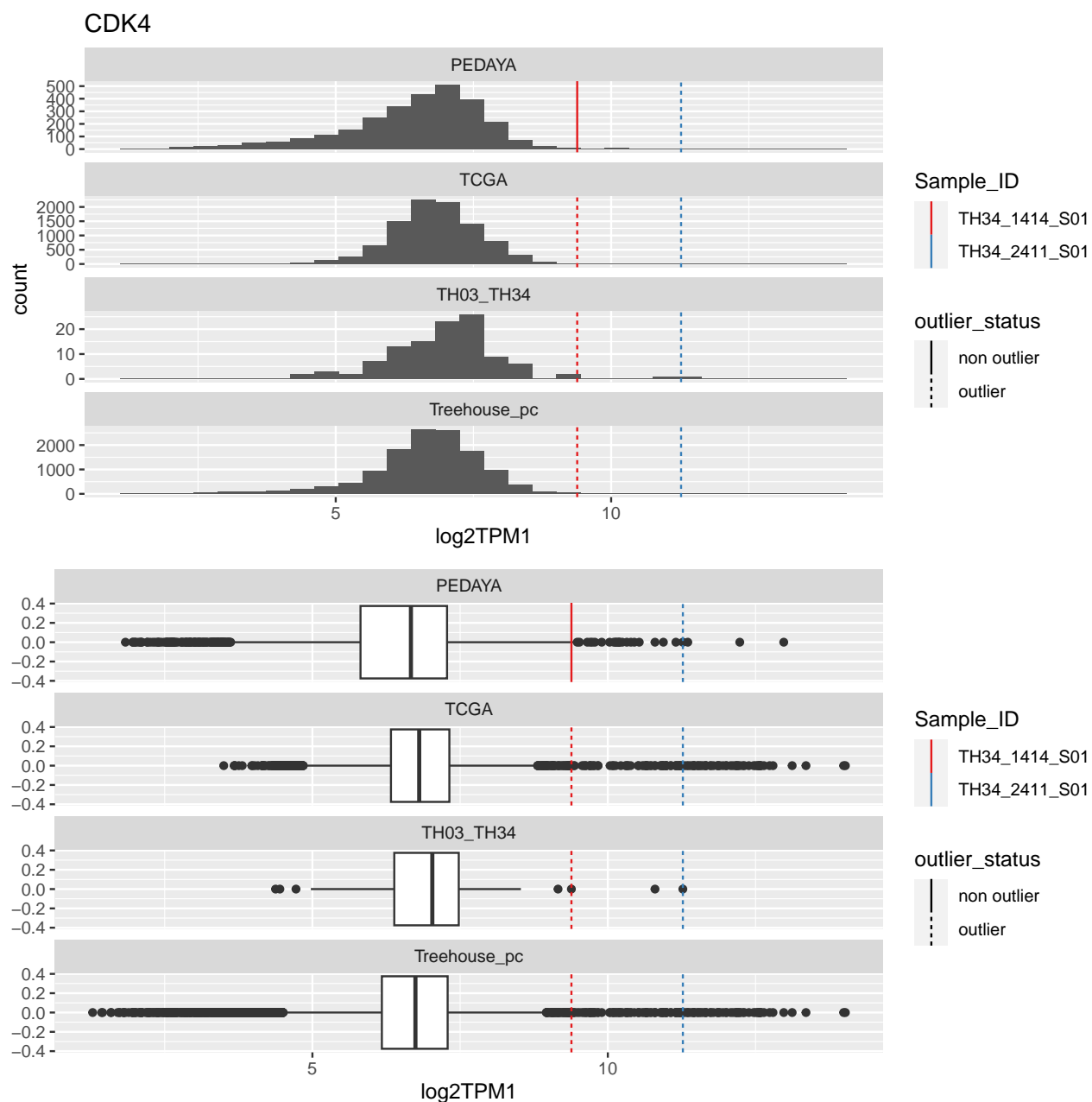
Sample_ID	gene	PEDAYA	TCGA	TH03_TH34	Treehouse_pc	log2TPM1
TH34_1452_S01	HDAC4	TRUE	TRUE	TRUE	TRUE	10.91

```
##  
## [[14]]
```



Sample_ID	gene	PEDAYA	TCGA	TH03_TH34	Treehouse_pc	log2TPM1
TH34_2292_S01	FLT4	TRUE	TRUE	TRUE	TRUE	7.679
TH34_2292_S01	FLT4	TRUE	TRUE	TRUE	TRUE	NA
TH34_1415_S01	FLT4	FALSE	TRUE	TRUE	TRUE	6.644
TH34_2666_S01	FLT4	FALSE	TRUE	FALSE	TRUE	5.512
TH34_2666_S01	FLT4	FALSE	TRUE	FALSE	TRUE	NA

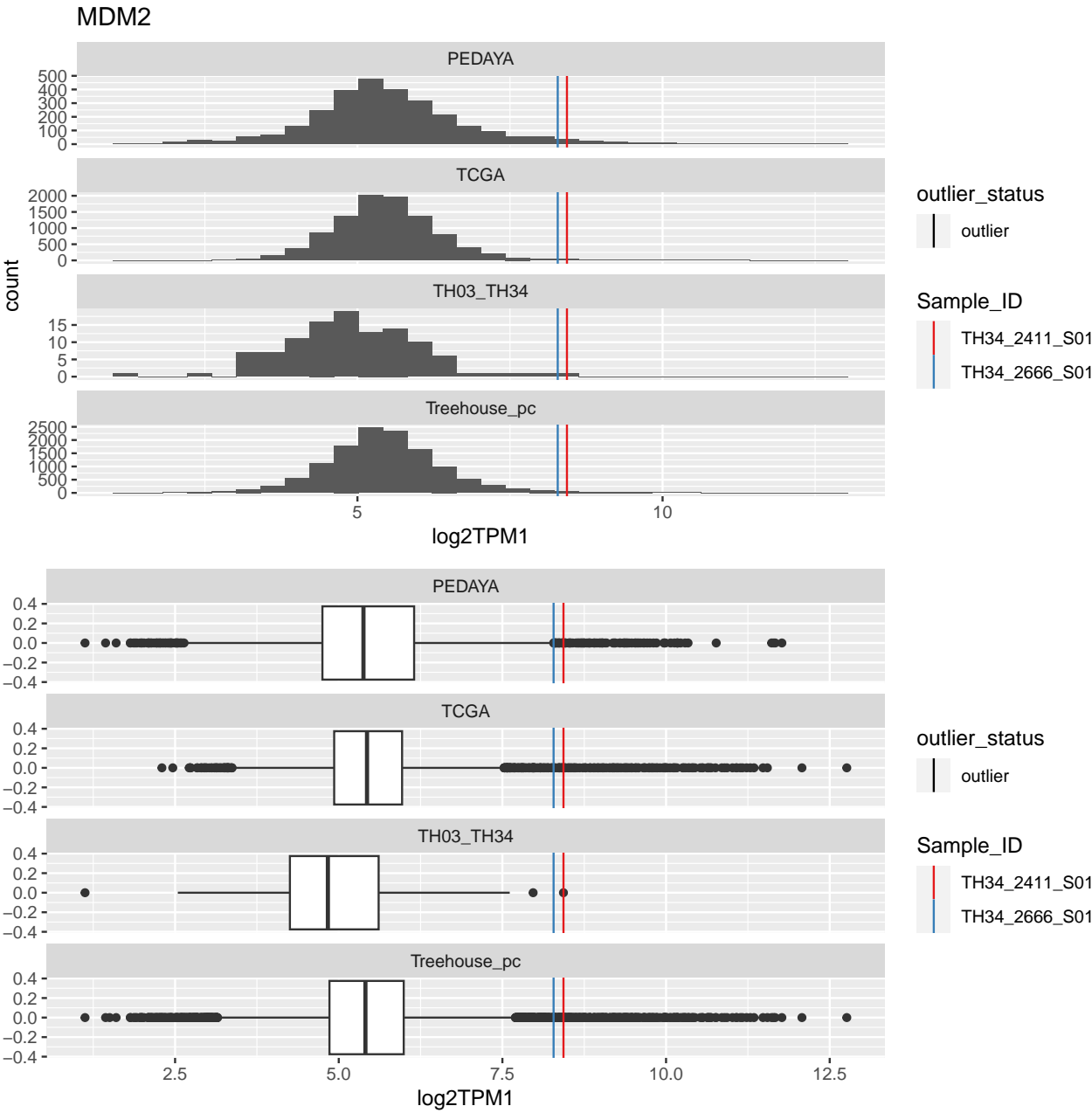
```
##  
## [[15]]
```



Sample_ID	gene	PEDAYA	TCGA	TH03_TH34	Treehouse_pc	log2TPM1
TH34_2411_S01	CDK4	TRUE	TRUE	TRUE	TRUE	11.270
TH34_1414_S01	CDK4	FALSE	TRUE	TRUE	TRUE	9.384

##

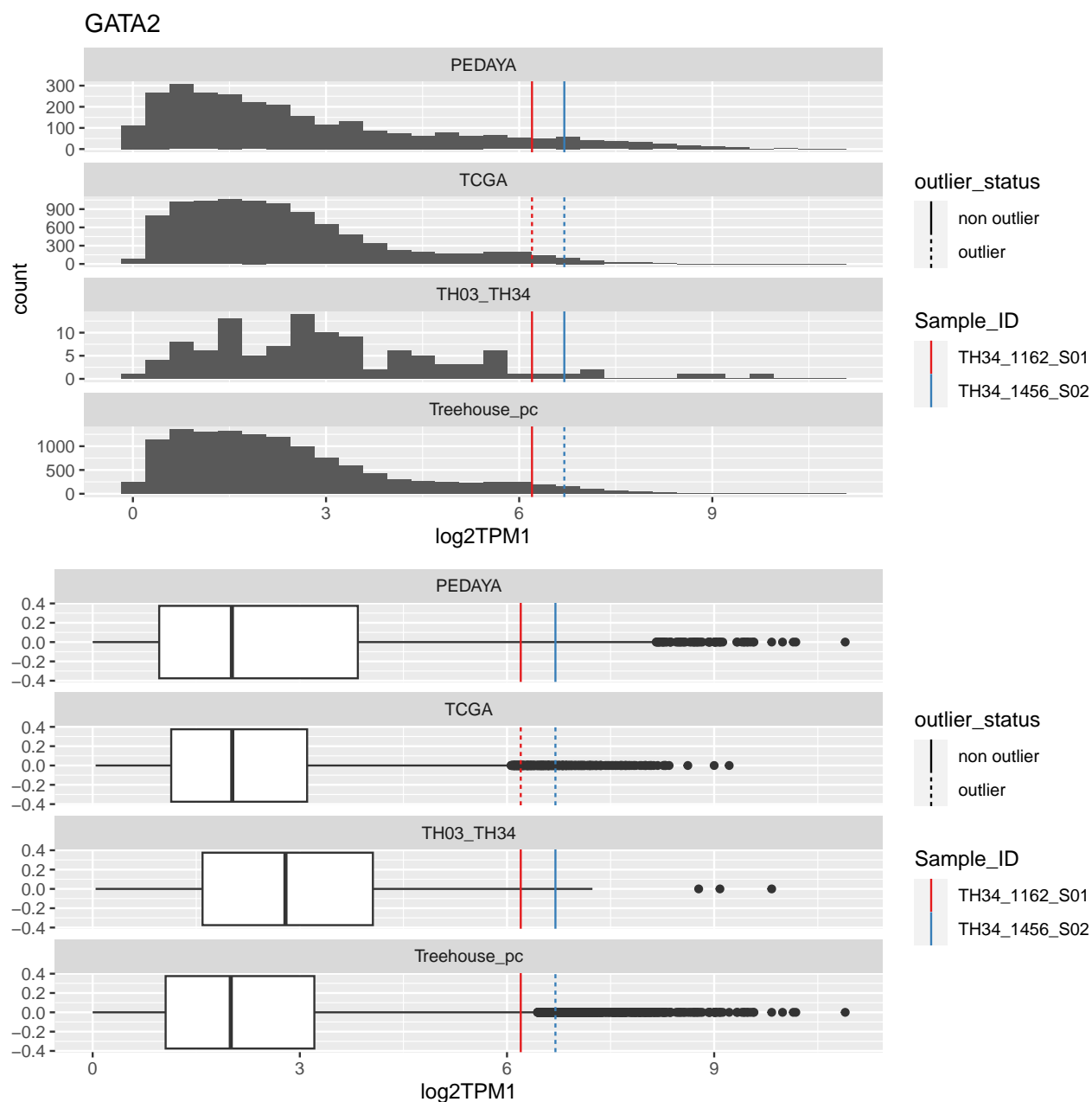
## [[16]]



Sample_ID	gene	PEDAYA	TCGA	TH03_TH34	Treehouse_pc	log2TPM1
TH34_2411_S01	MDM2	TRUE	TRUE	TRUE	TRUE	8.433
TH34_2666_S01	MDM2	TRUE	TRUE	TRUE	TRUE	8.282
TH34_2666_S01	MDM2	TRUE	TRUE	TRUE	TRUE	NA



```
##  
## [[17]]
```



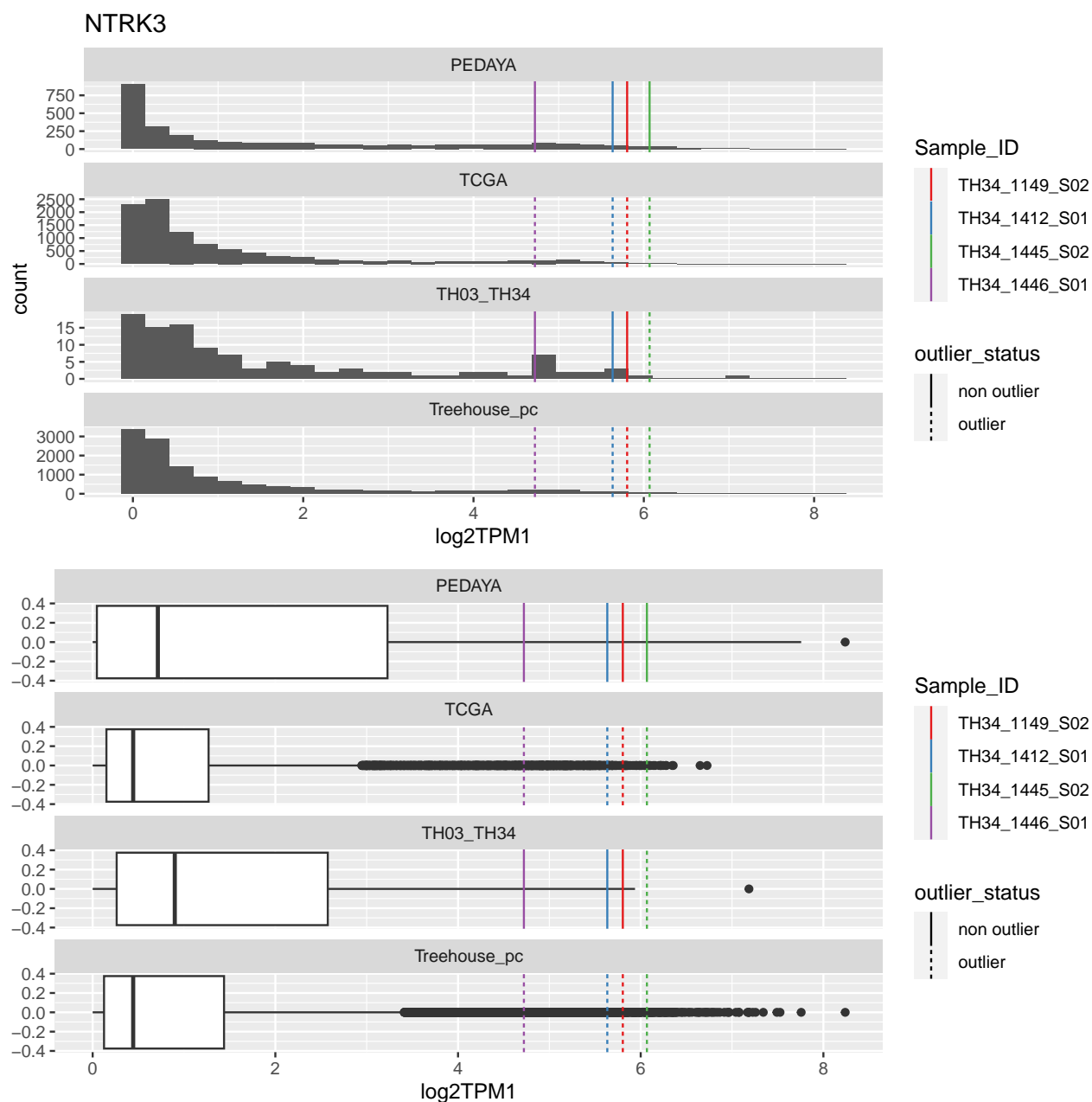
Sample_ID	gene	PEDAYA	TCGA	TH03_TH34	Treehouse_pc	log2TPM1
TH34_1162_S01	GATA2	FALSE	TRUE	FALSE	FALSE	6.199
TH34_1456_S02	GATA2	FALSE	TRUE	FALSE	TRUE	6.701
TH34_1456_S02	GATA2	FALSE	TRUE	FALSE	TRUE	NA

##

## [[18]]

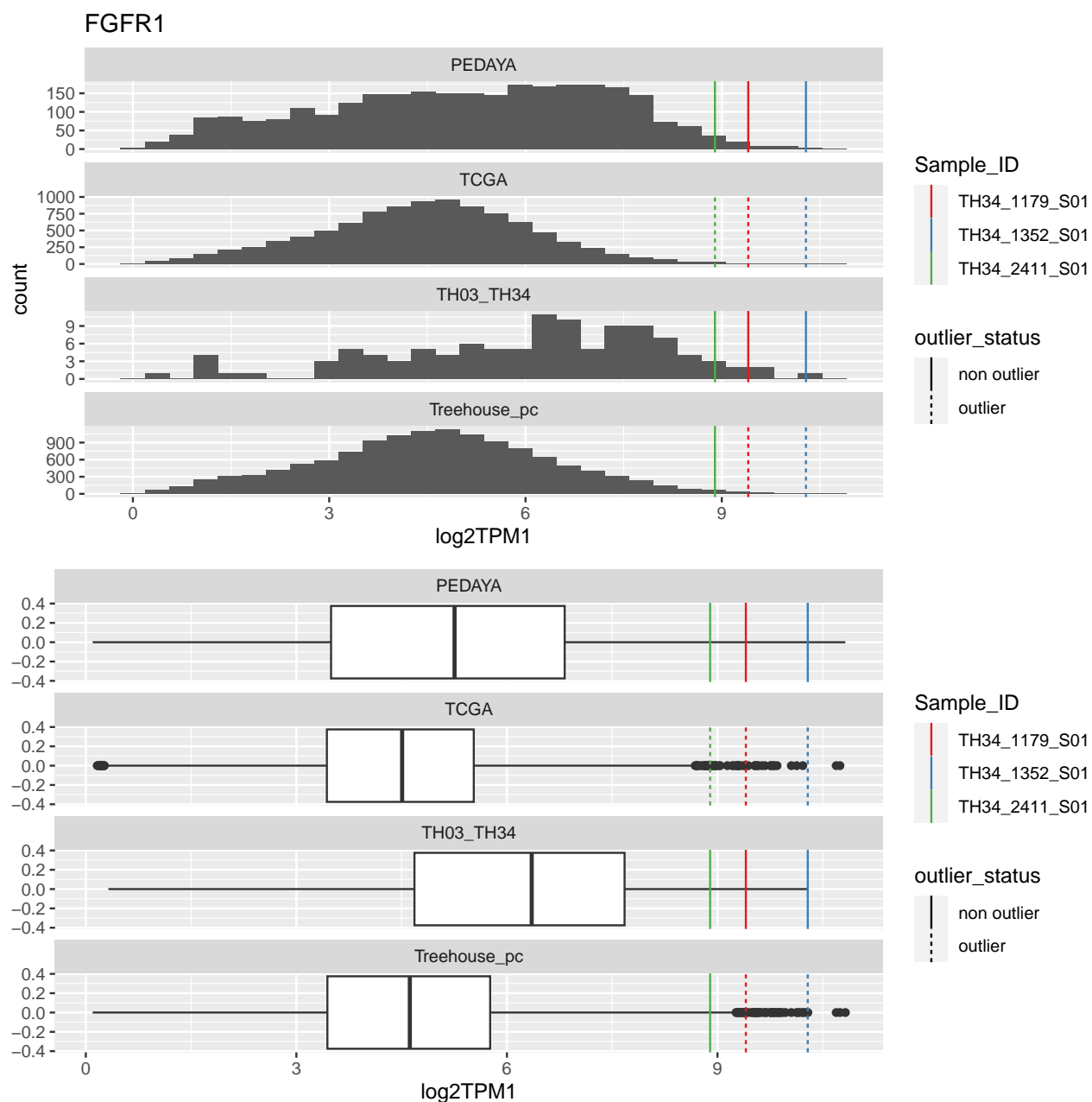


```
##  
## [[19]]
```



Sample_ID	gene	PEDAYA	TCGA	TH03_TH34	Treehouse_pc	log2TPM1
TH34_1149_S02	NTRK3	FALSE	TRUE	FALSE	TRUE	5.804
TH34_1412_S01	NTRK3	FALSE	TRUE	FALSE	TRUE	5.634
TH34_1445_S02	NTRK3	FALSE	TRUE	TRUE	TRUE	6.068
TH34_1445_S02	NTRK3	FALSE	TRUE	TRUE	TRUE	NA
TH34_1446_S01	NTRK3	FALSE	TRUE	FALSE	TRUE	4.721

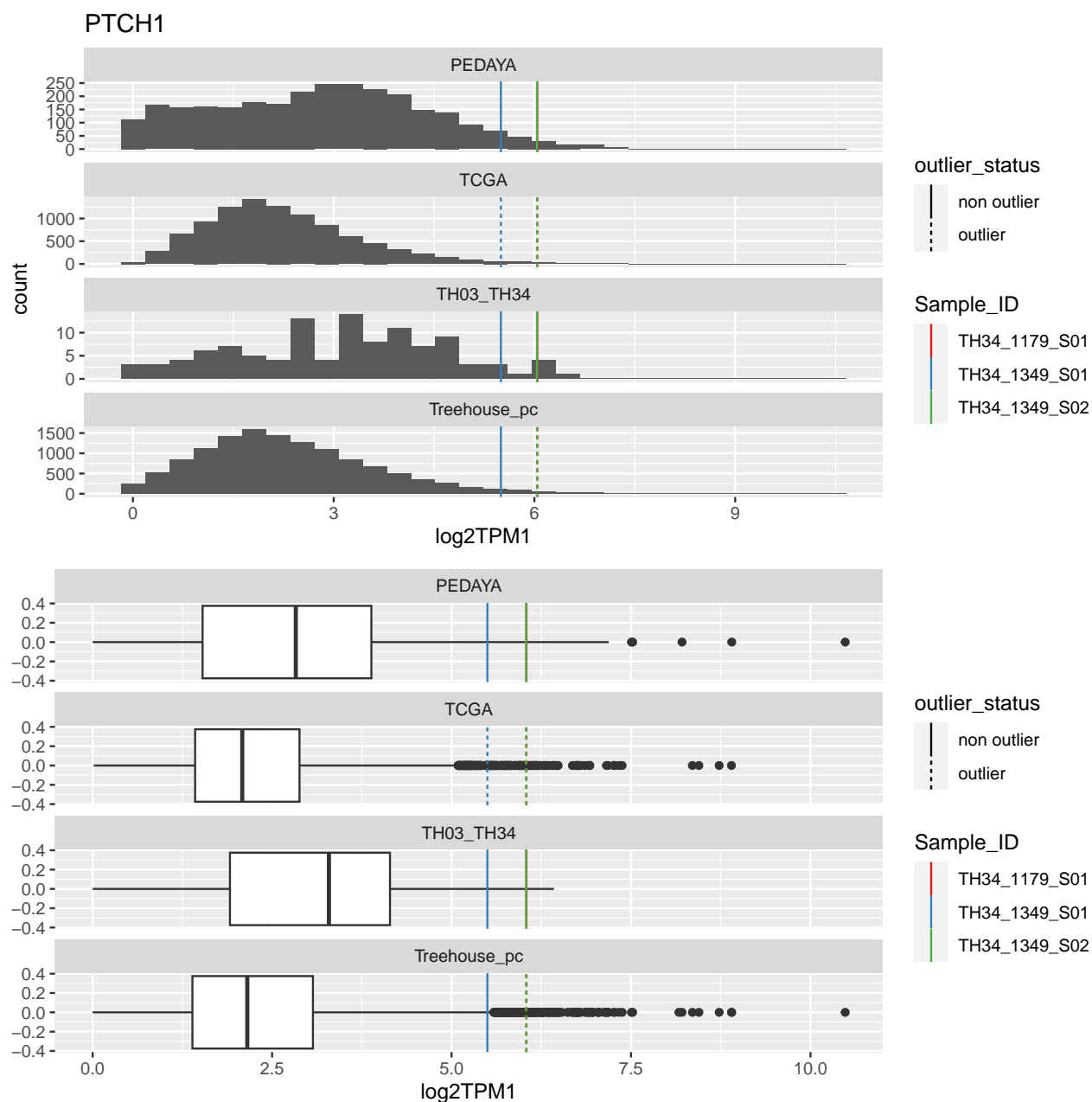
```
##  
## [[20]]
```



Sample_ID	gene	PEDAYA	TCGA	TH03_TH34	Treehouse_pc	log2TPM1
TH34_1179_S01	FGFR1	FALSE	TRUE	FALSE	TRUE	9.404
TH34_1352_S01	FGFR1	FALSE	TRUE	FALSE	TRUE	10.286
TH34_2411_S01	FGFR1	FALSE	TRUE	FALSE	FALSE	8.895



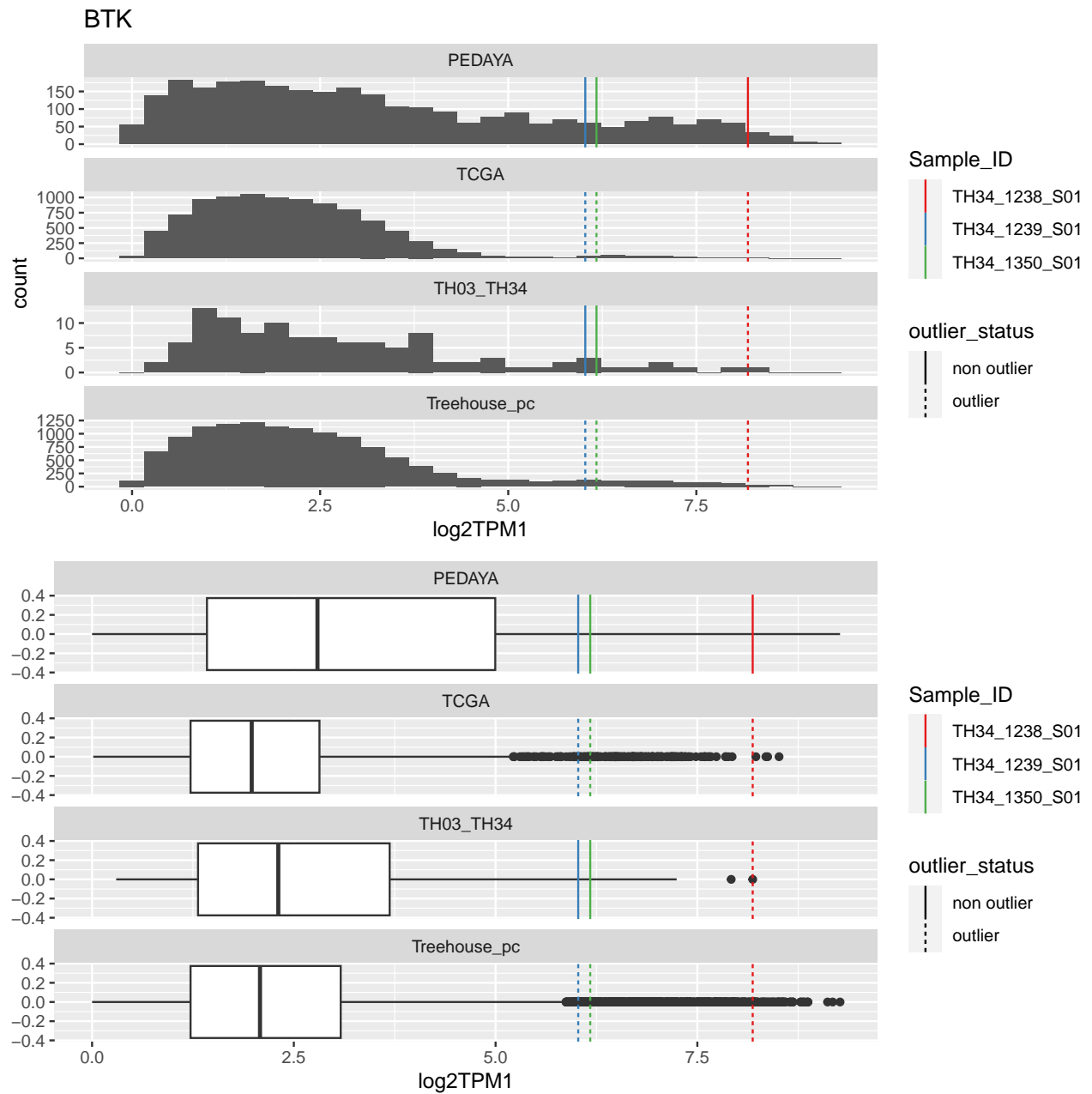
```
##  
## [[21]]
```



Sample_ID	gene	PEDAYA	TCGA	TH03_TH34	Treehouse_pc	log2TPM1
TH34_1179_S01	PTCH1	FALSE	TRUE	FALSE	TRUE	6.043
TH34_1349_S01	PTCH1	FALSE	TRUE	FALSE	FALSE	5.499
TH34_1349_S02	PTCH1	FALSE	TRUE	FALSE	TRUE	6.038

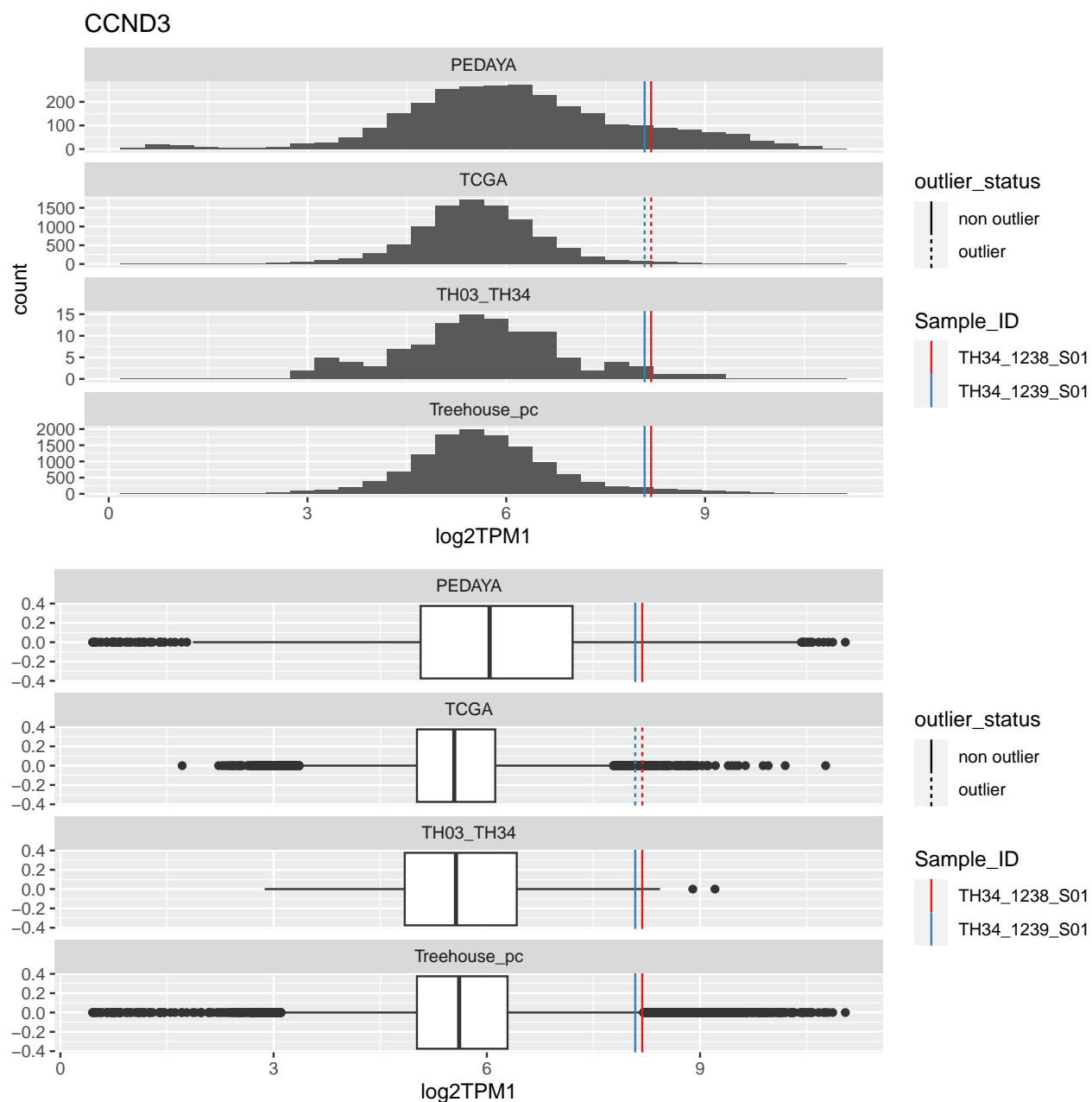
##

```
## [[22]]
```



Sample_ID	gene	PEDAYA	TCGA	TH03_TH34	Treehouse_pc	log2TPM1
TH34_1238_S01	BTK	FALSE	TRUE	TRUE	TRUE	8.186
TH34_1239_S01	BTK	FALSE	TRUE	FALSE	TRUE	6.024
TH34_1350_S01	BTK	FALSE	TRUE	FALSE	TRUE	6.173

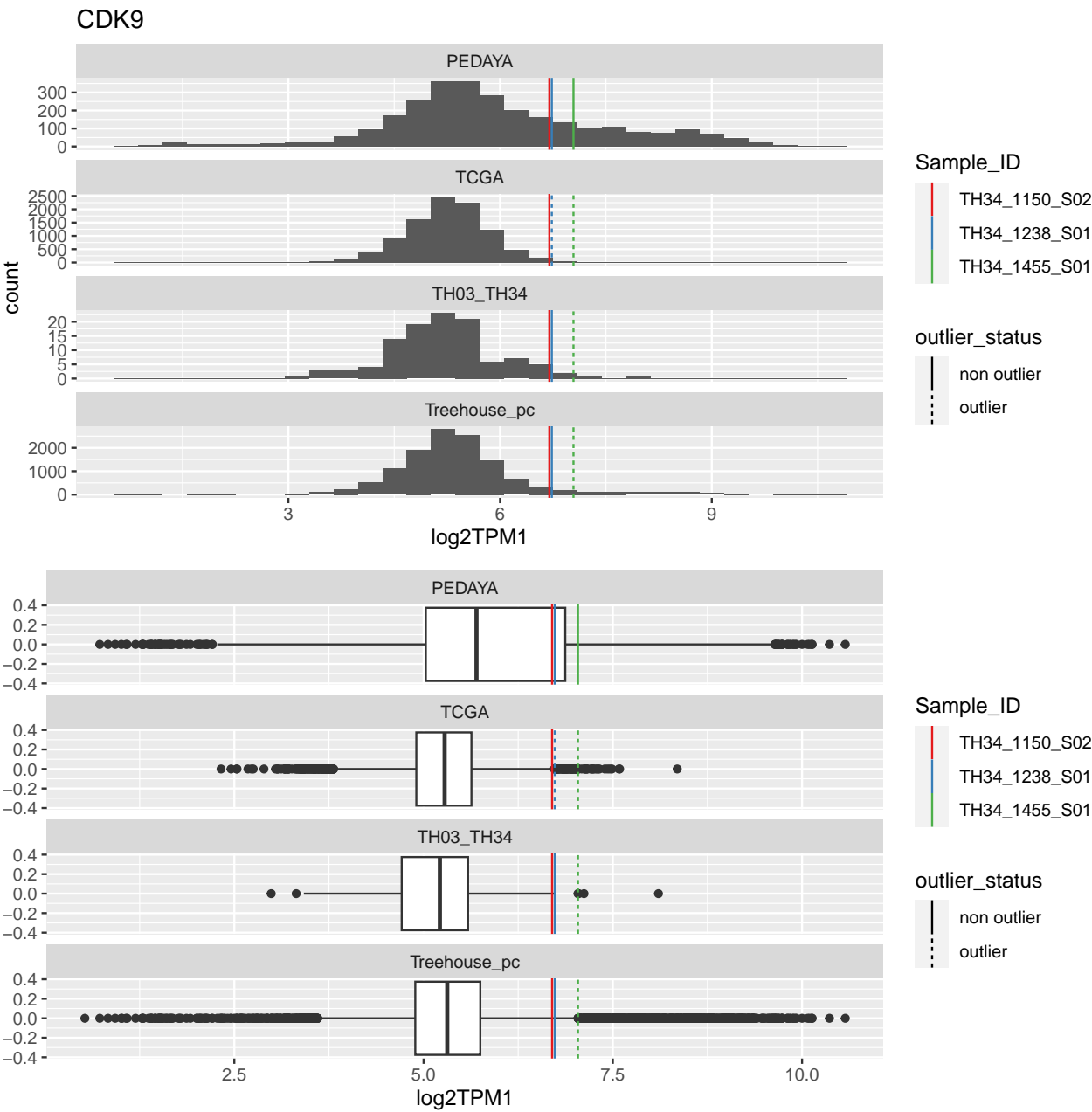
```
##  
## [[23]]
```



Sample_ID	gene	PEDAYA	TCGA	TH03_TH34	Treehouse_pc	log2TPM1
TH34_1238_S01	CCND3	FALSE	TRUE	FALSE	FALSE	8.186
TH34_1239_S01	CCND3	FALSE	TRUE	FALSE	FALSE	8.088

##

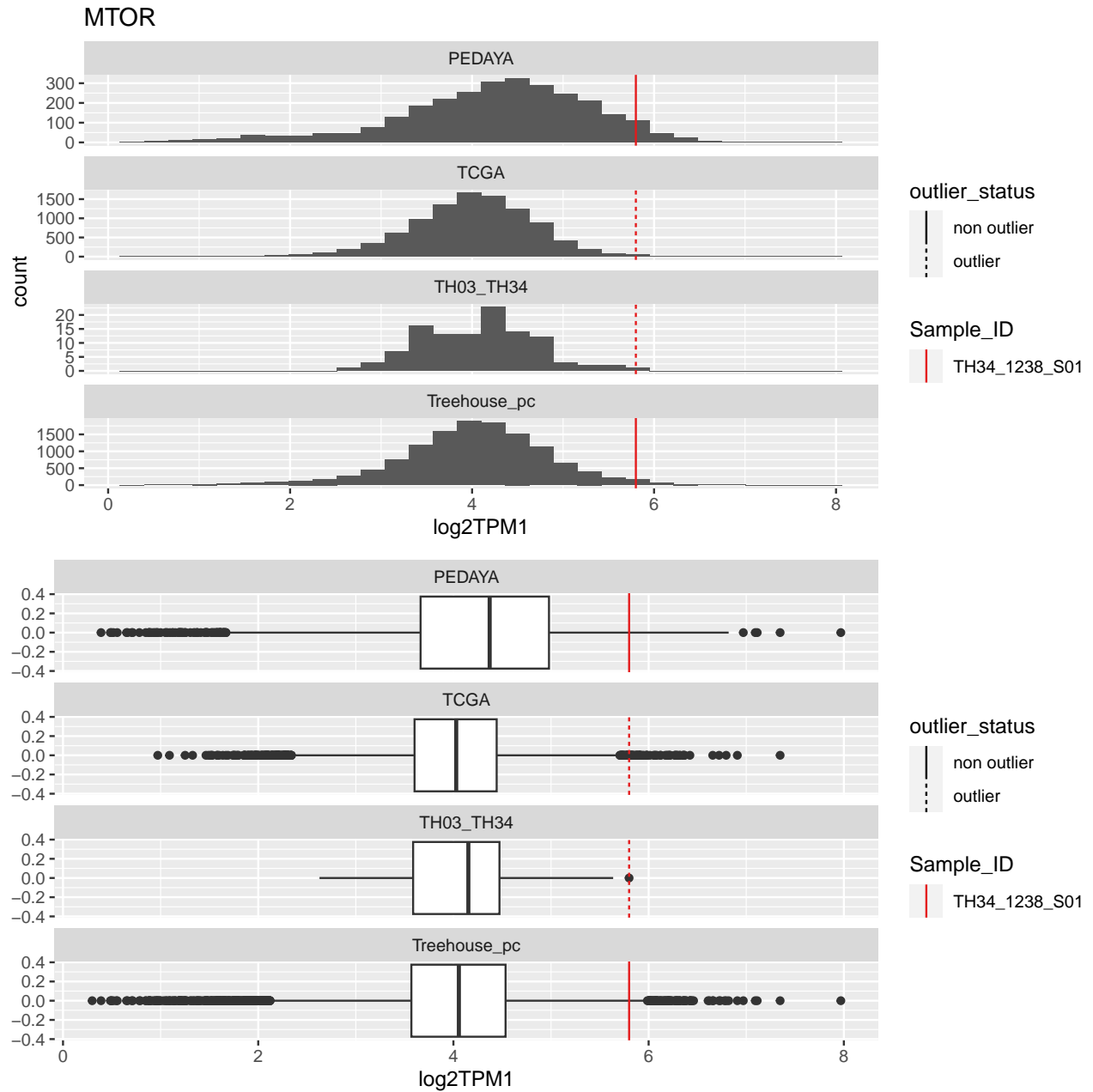
## [[24]]



Sample_ID	gene	PEDAYA	TCGA	TH03_TH34	Treehouse_pc	log2TPM1
TH34_1238_S01	CDK9	FALSE	TRUE	FALSE	FALSE	6.733
TH34_1455_S01	CDK9	FALSE	TRUE	TRUE	TRUE	7.040
TH34_1150_S02	CDK9	FALSE	FALSE	FALSE	FALSE	6.698

##

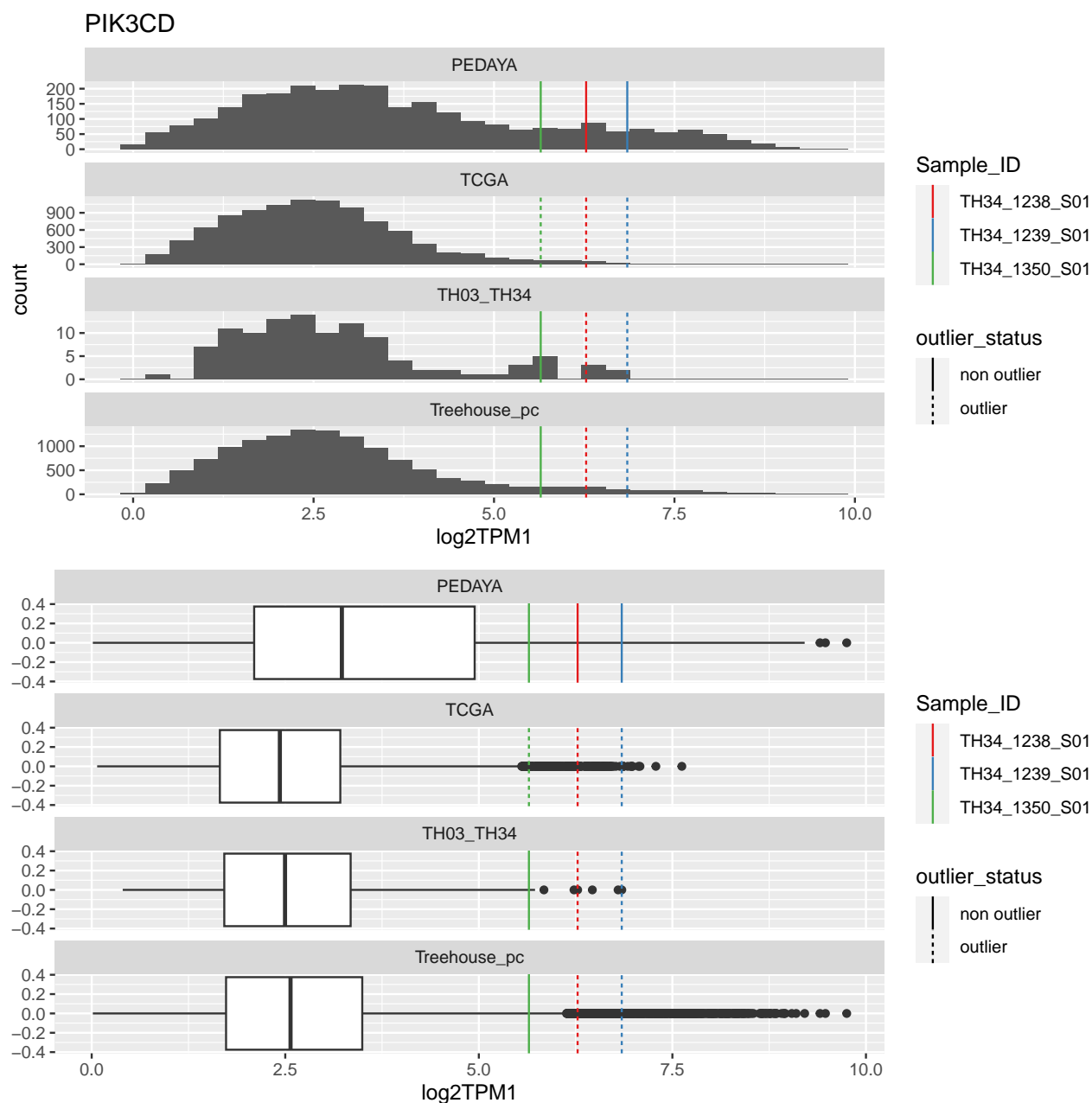
## [[25]]



Sample_ID	gene	PEDAYA	TCGA	TH03_TH34	Treehouse_pc	log2TPM1
TH34_1238_S01	MTOR	FALSE	TRUE	TRUE	FALSE	5.8

```
##  
## [[26]]
```

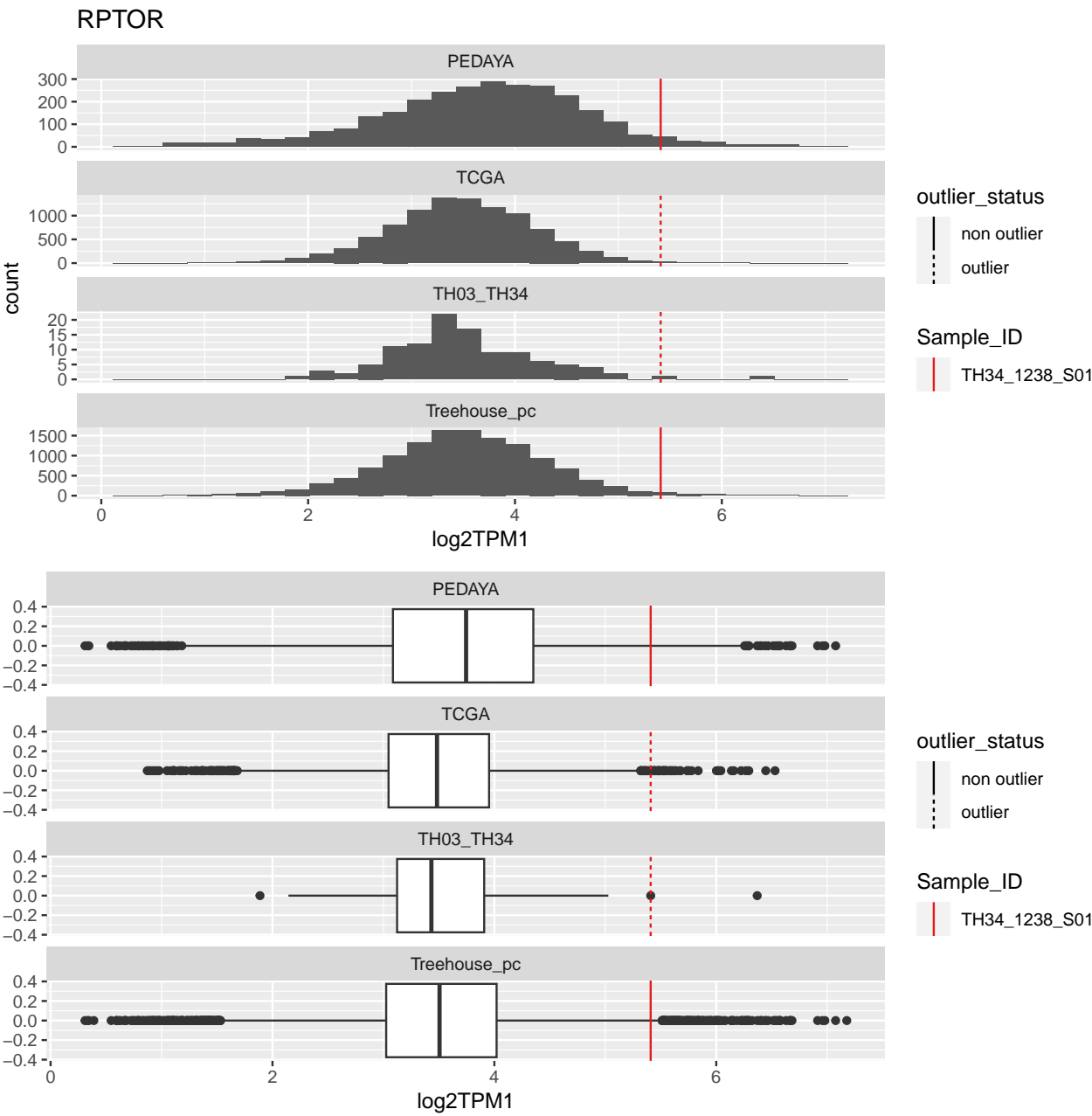




Sample_ID	gene	PEDAYA	TCGA	TH03_TH34	Treehouse_pc	log2TPM1
TH34_1238_S01	PIK3CD	FALSE	TRUE	TRUE	TRUE	6.276
TH34_1239_S01	PIK3CD	FALSE	TRUE	TRUE	TRUE	6.846
TH34_1350_S01	PIK3CD	FALSE	TRUE	FALSE	FALSE	5.647

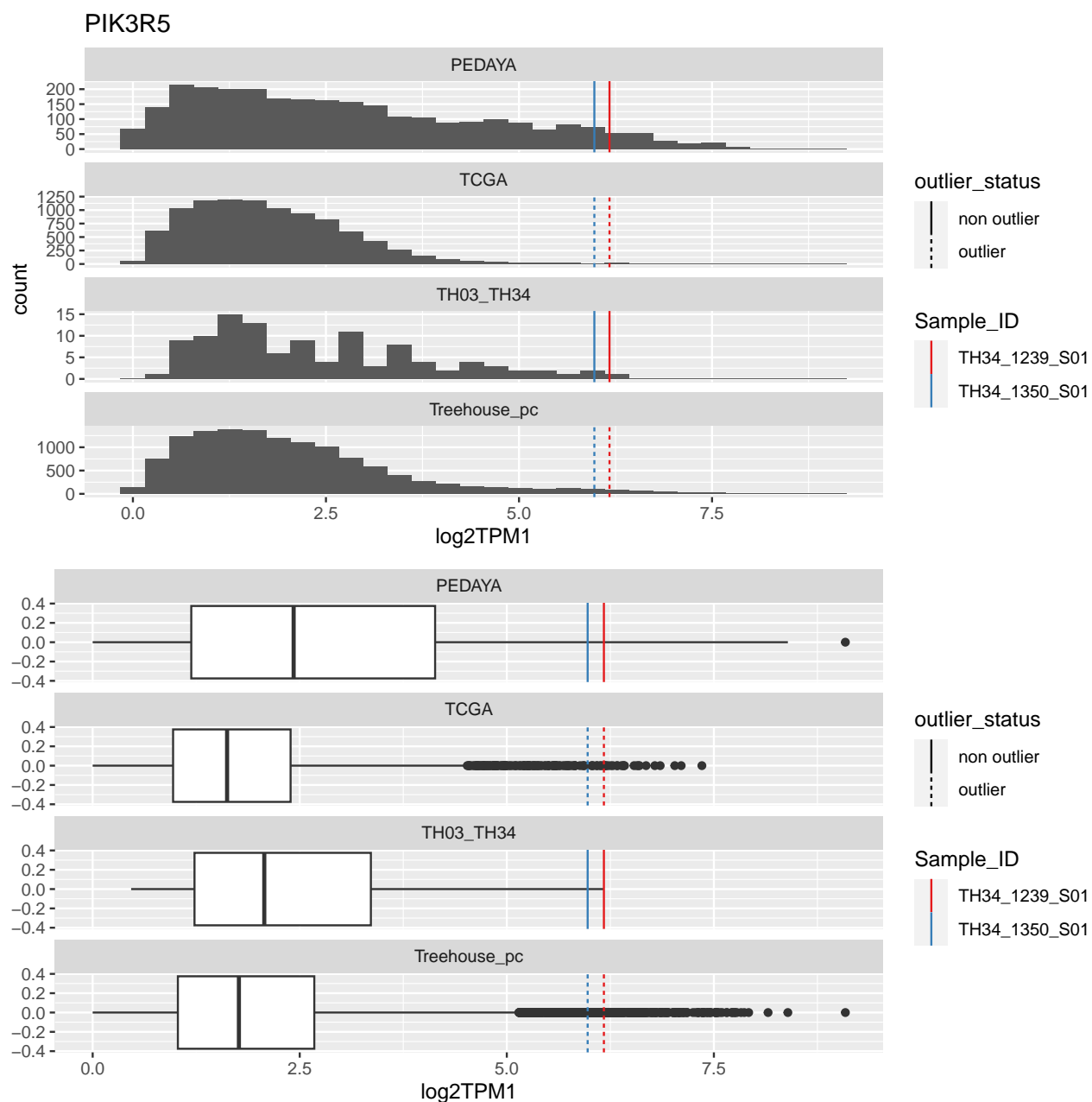
##

## [[27]]



Sample_ID	gene	PEDAYA	TCGA	TH03_TH34	Treehouse_pc	log2TPM1
TH34_1238_S01	RPTOR	FALSE	TRUE	TRUE	FALSE	5.409

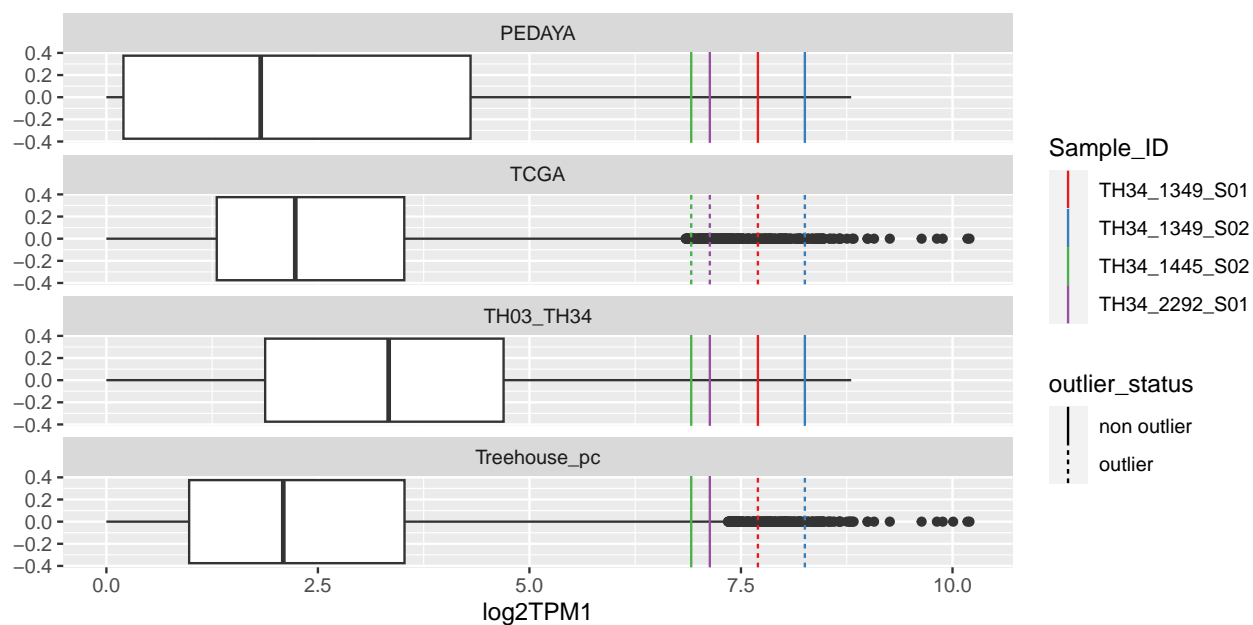
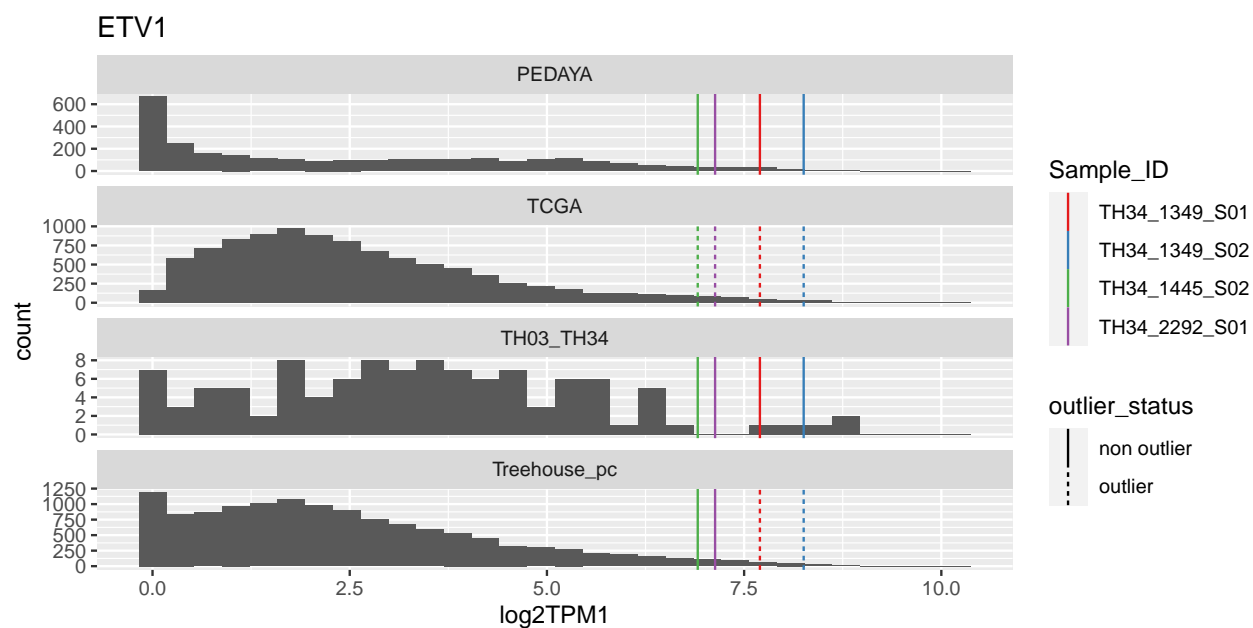
```
##  
## [[28]]
```



Sample_ID	gene	PEDAYA	TCGA	TH03_TH34	Treehouse_pc	log2TPM1
TH34_1239_S01	PIK3R5	FALSE	TRUE	FALSE	TRUE	6.172
TH34_1350_S01	PIK3R5	FALSE	TRUE	FALSE	TRUE	5.976

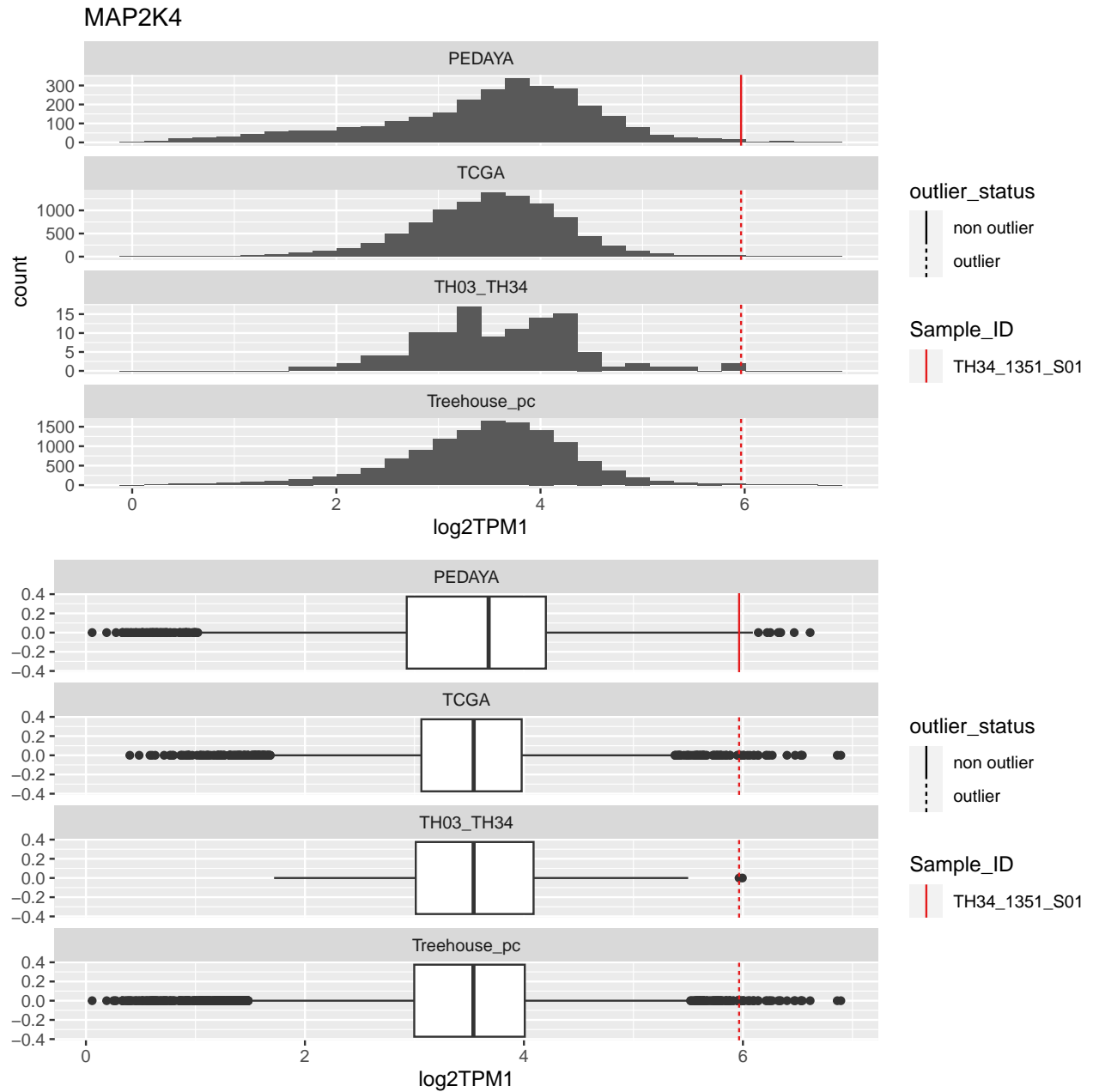
##

## [[29]]



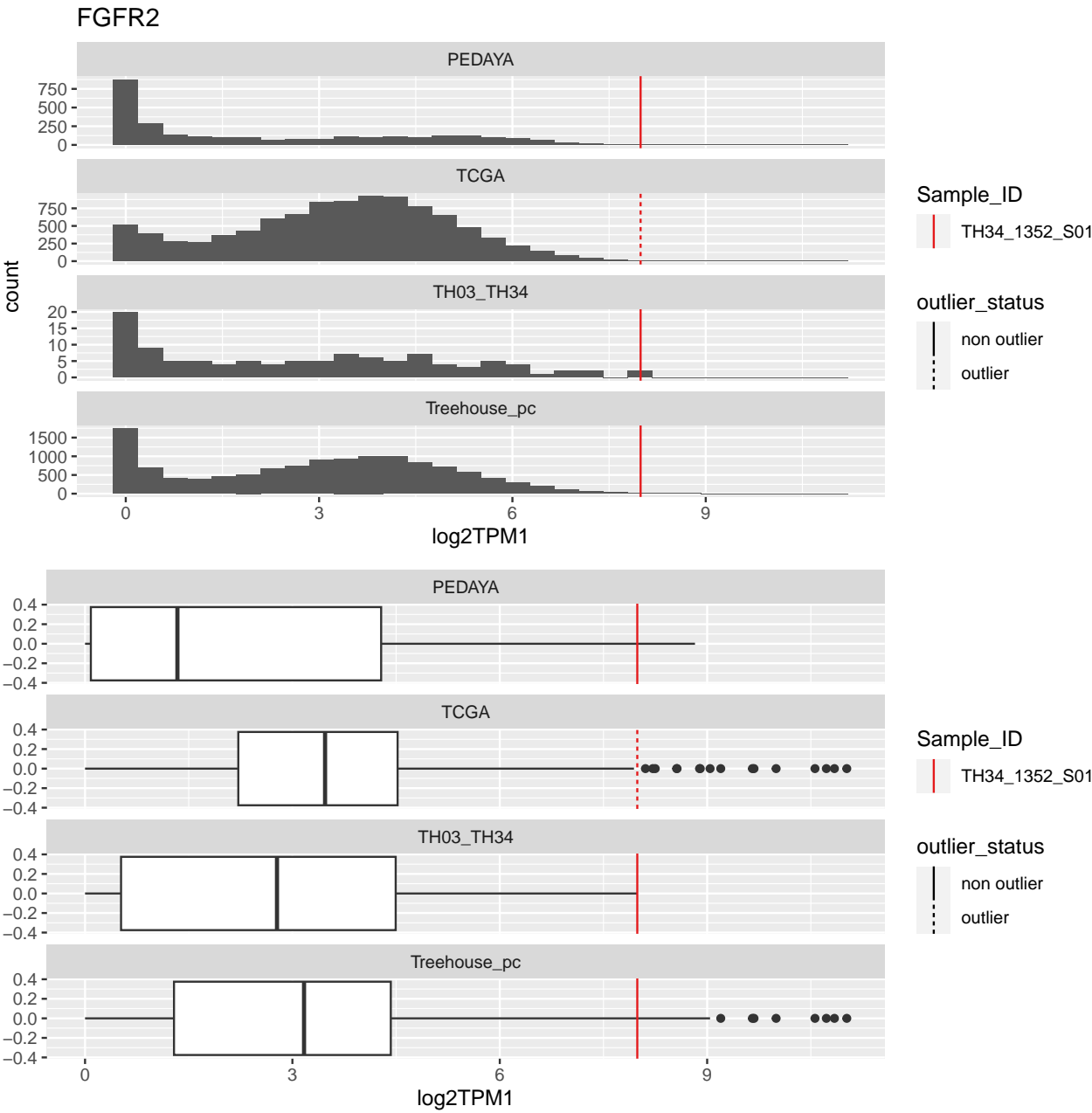
Sample_ID	gene	PEDAYA	TCGA	TH03_TH34	Treehouse_pc	log2TPM1
TH34_1349_S01	ETV1	FALSE	TRUE	FALSE	TRUE	7.700
TH34_1349_S02	ETV1	FALSE	TRUE	FALSE	TRUE	8.255
TH34_1445_S02	ETV1	FALSE	TRUE	FALSE	FALSE	6.912
TH34_1445_S02	ETV1	FALSE	TRUE	FALSE	FALSE	NA
TH34_2292_S01	ETV1	FALSE	TRUE	FALSE	FALSE	7.132
TH34_2292_S01	ETV1	FALSE	TRUE	FALSE	FALSE	NA

```
##  
## [[30]]
```



Sample_ID	gene	PEDAYA	TCGA	TH03_TH34	Treehouse_pc	log2TPM1
TH34_1351_S01	MAP2K4	FALSE	TRUE	TRUE	TRUE	5.965

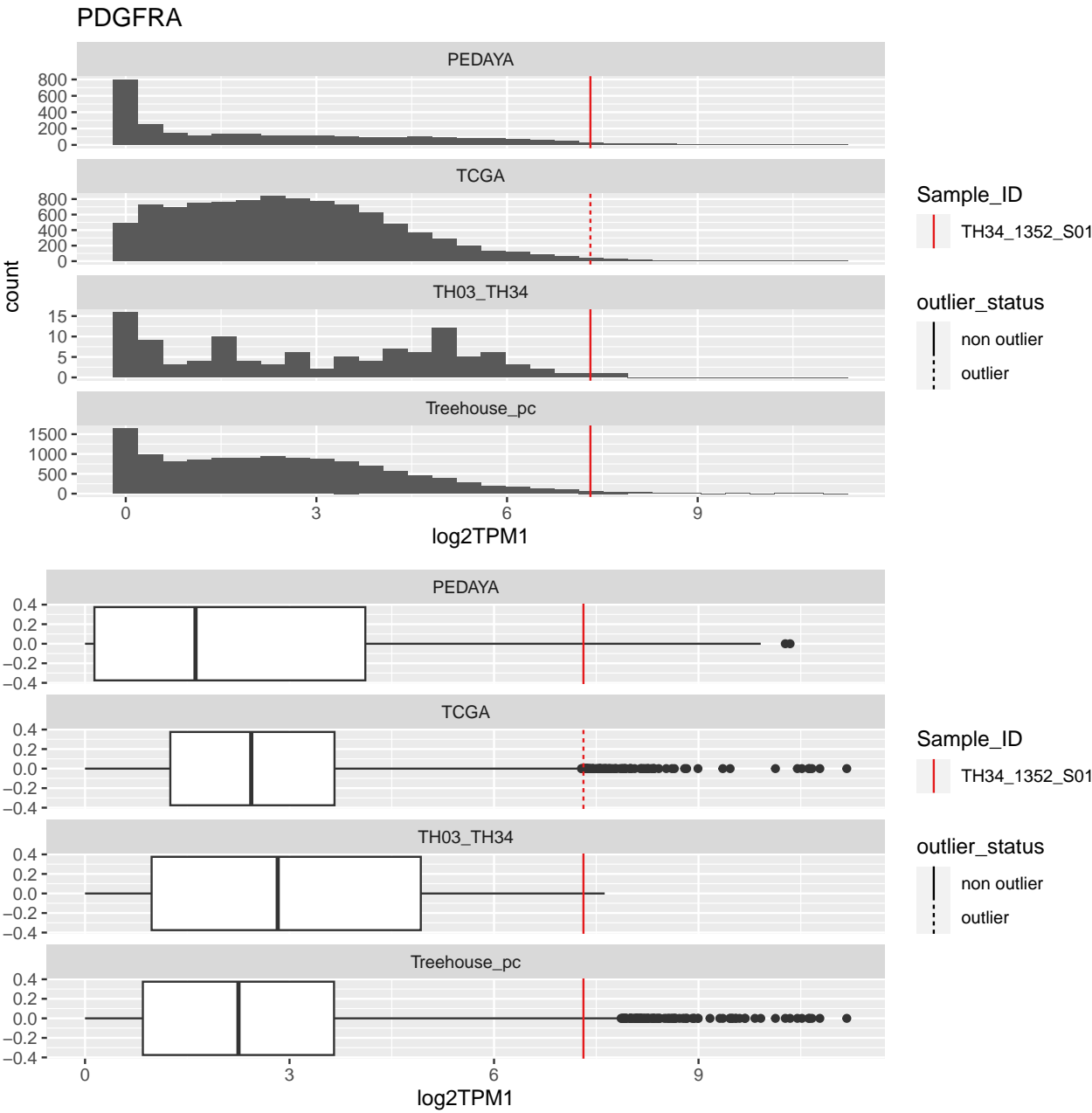
##  
## [[31]]



Sample_ID	gene	PEDAYA	TCGA	TH03_TH34	Treehouse_pc	log2TPM1
TH34_1352_S01	FGFR2	FALSE	TRUE	FALSE	FALSE	7.989



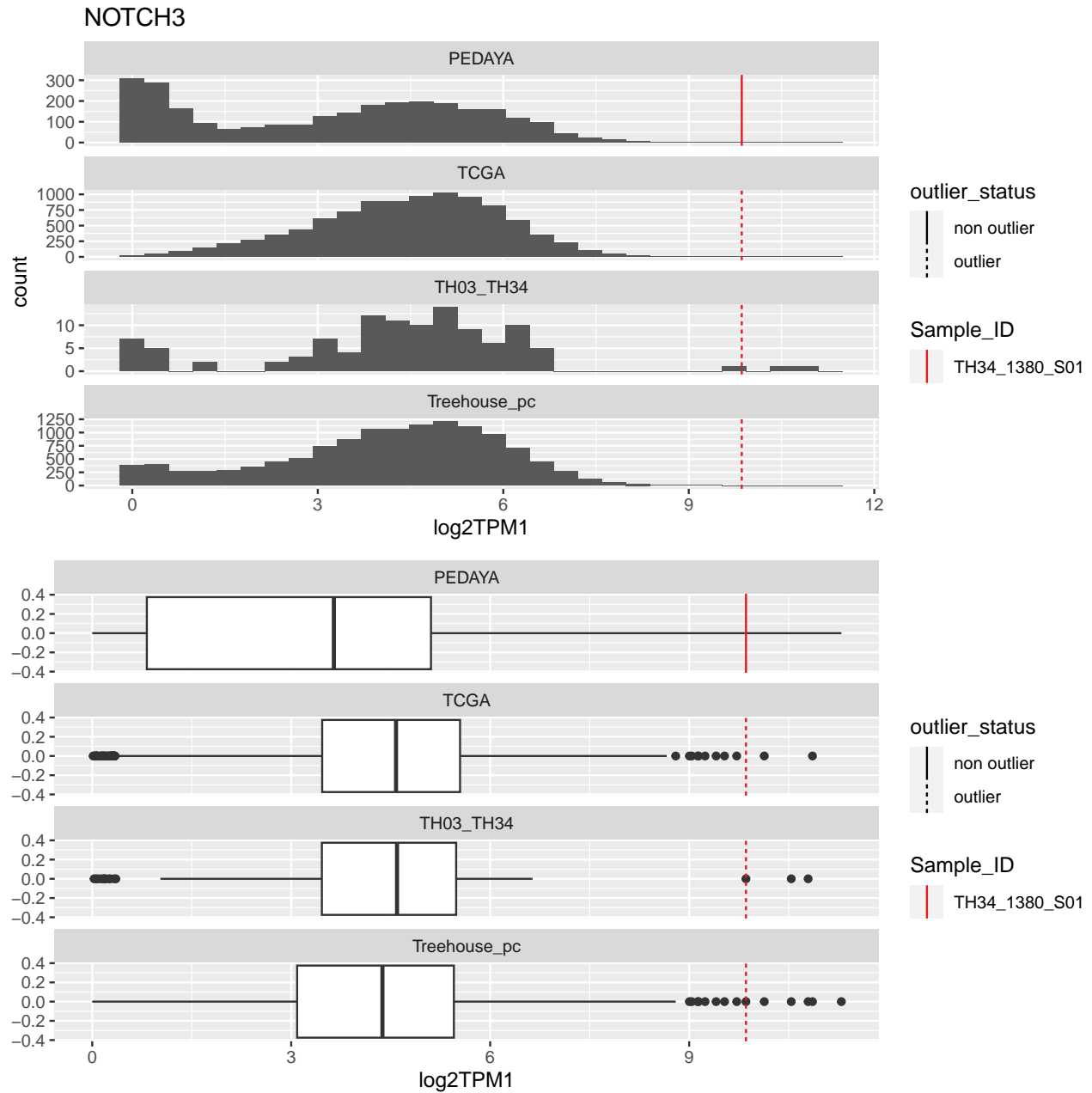
##  
## [[32]]



Sample_ID	gene	PEDAYA	TCGA	TH03_TH34	Treehouse_pc	log2TPM1
TH34_1352_S01	PDGFRA	FALSE	TRUE	FALSE	FALSE	7.312

##

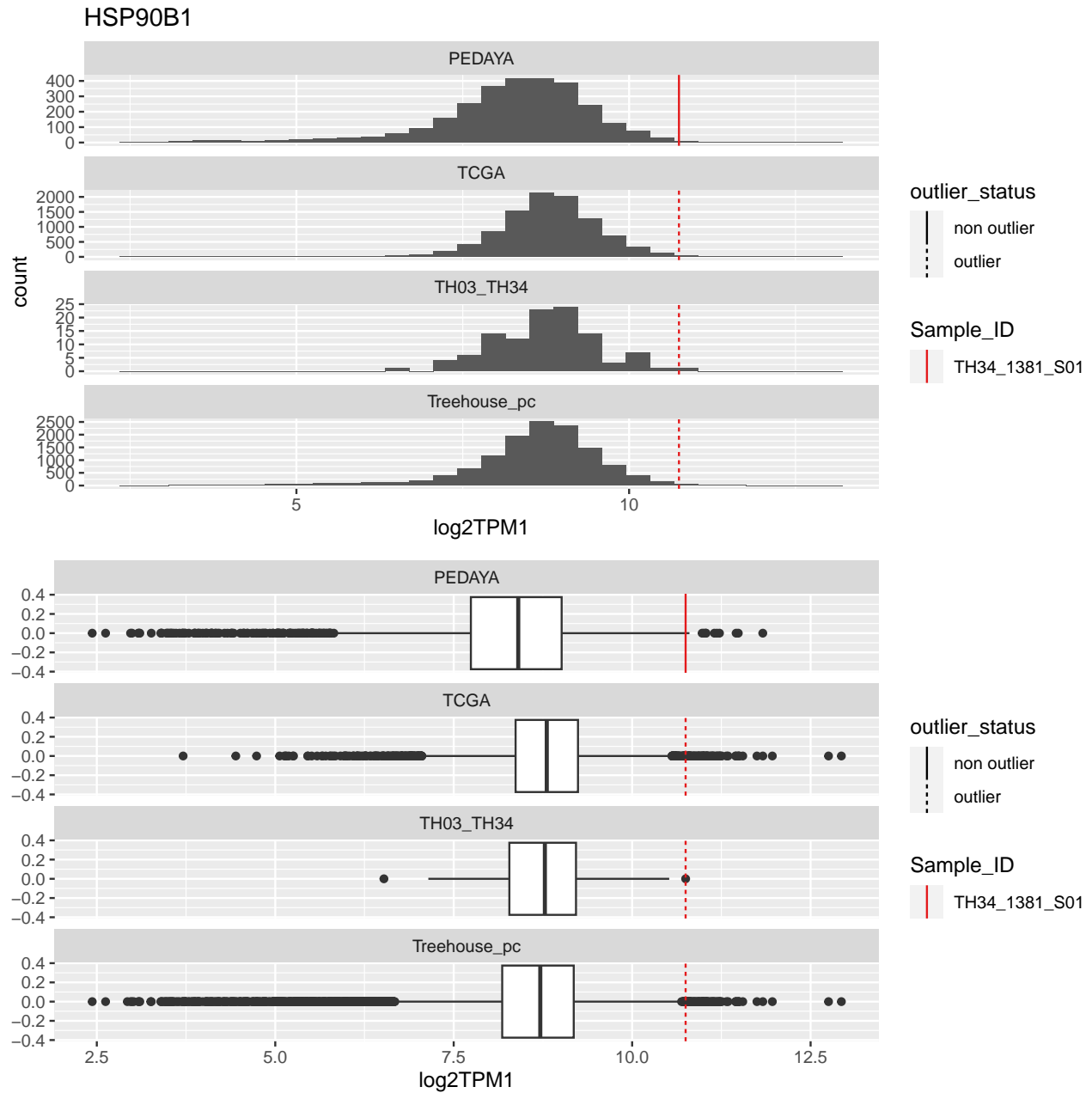
## [[33]]



Sample_ID	gene	PEDAYA	TCGA	TH03_TH34	Treehouse_pc	log2TPM1
TH34_1380_S01	NOTCH3	FALSE	TRUE	TRUE	TRUE	9.857

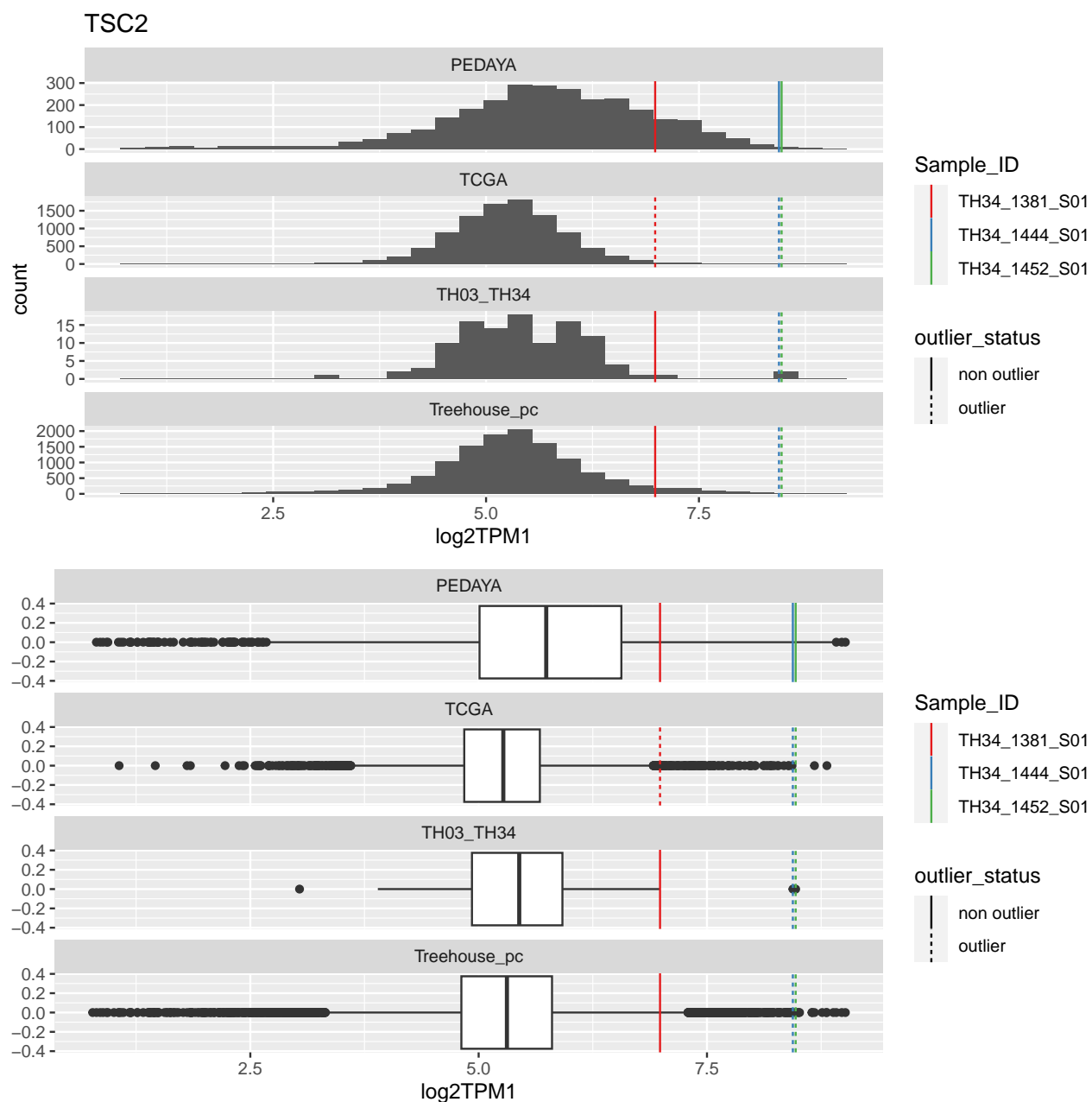
##

## [[34]]



Sample_ID	gene	PEDAYA	TCGA	TH03_TH34	Treehouse_pc	log2TPM1
TH34_1381_S01	HSP90B1	FALSE	TRUE	TRUE	TRUE	10.749

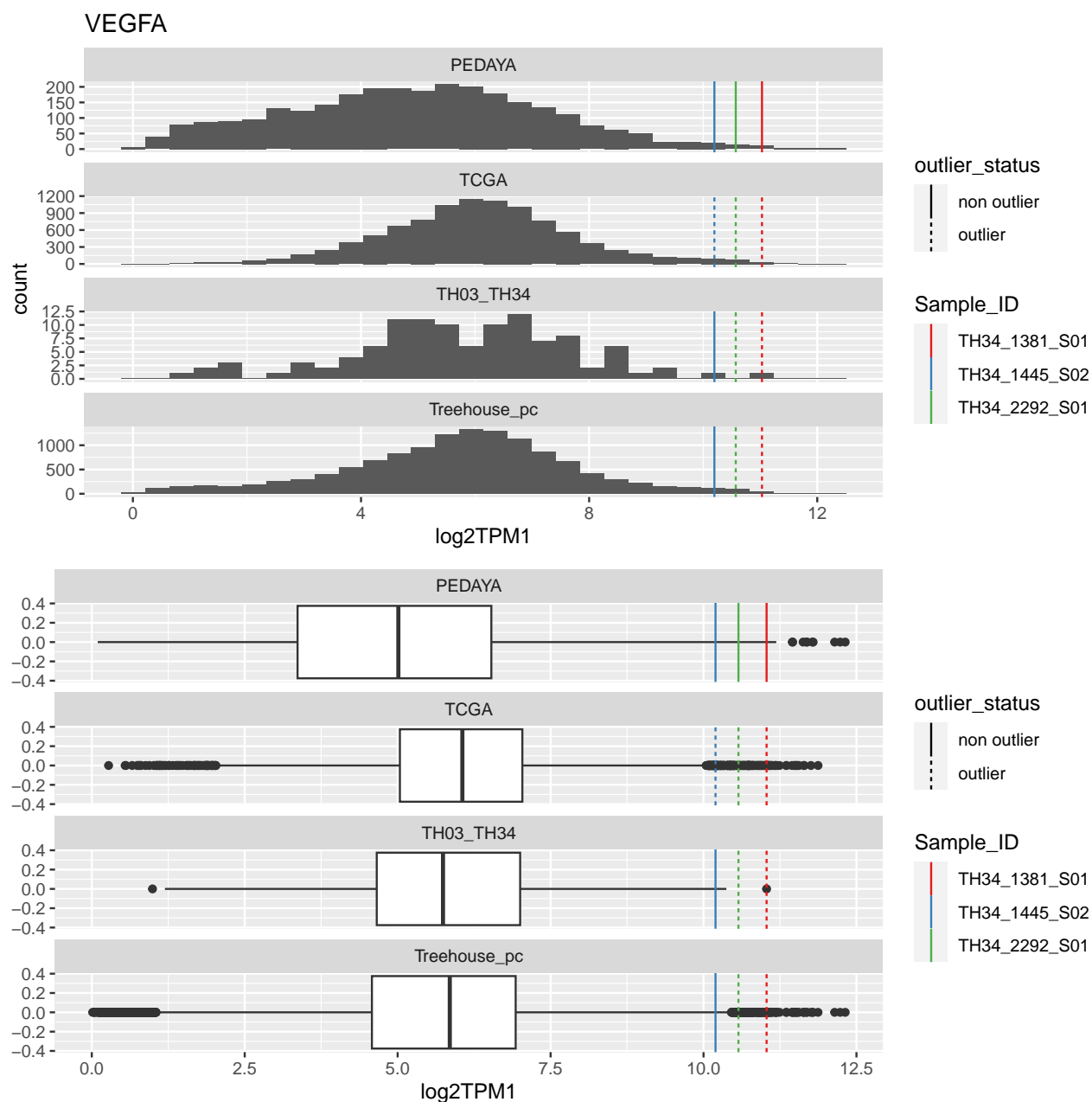
```
##  
## [[35]]
```



Sample_ID	gene	PEDAYA	TCGA	TH03_TH34	Treehouse_pc	log2TPM1
TH34_1381_S01	TSC2	FALSE	TRUE	FALSE	FALSE	6.986
TH34_1444_S01	TSC2	FALSE	TRUE	TRUE	TRUE	8.440
TH34_1452_S01	TSC2	FALSE	TRUE	TRUE	TRUE	8.471

##

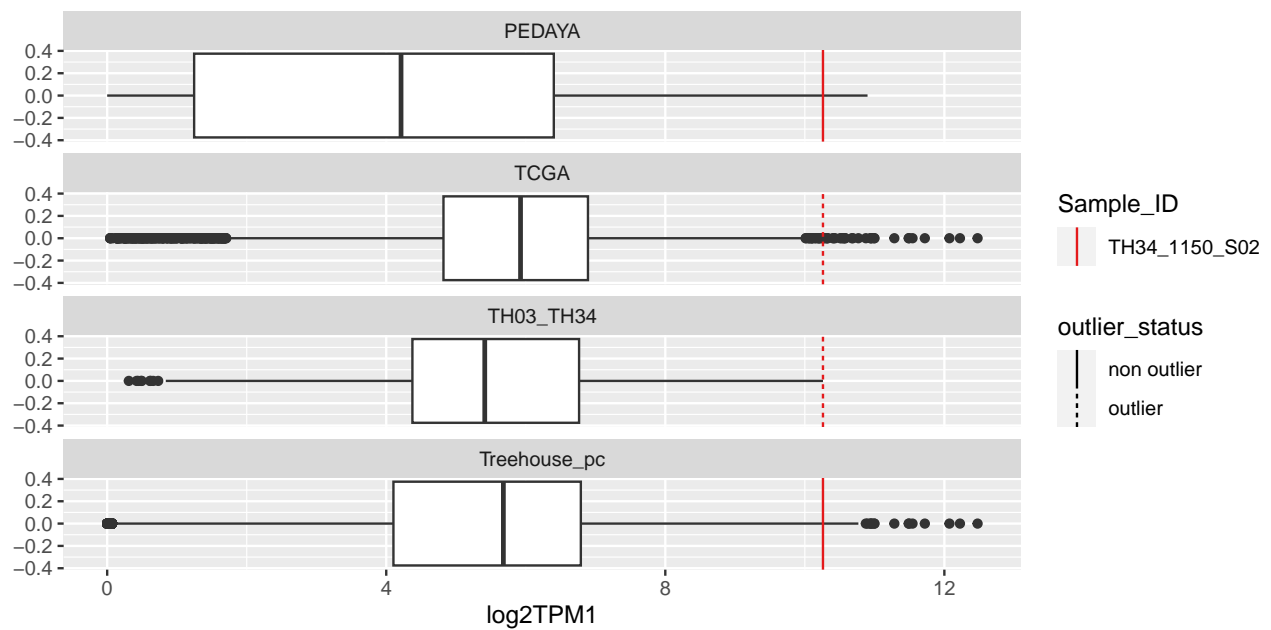
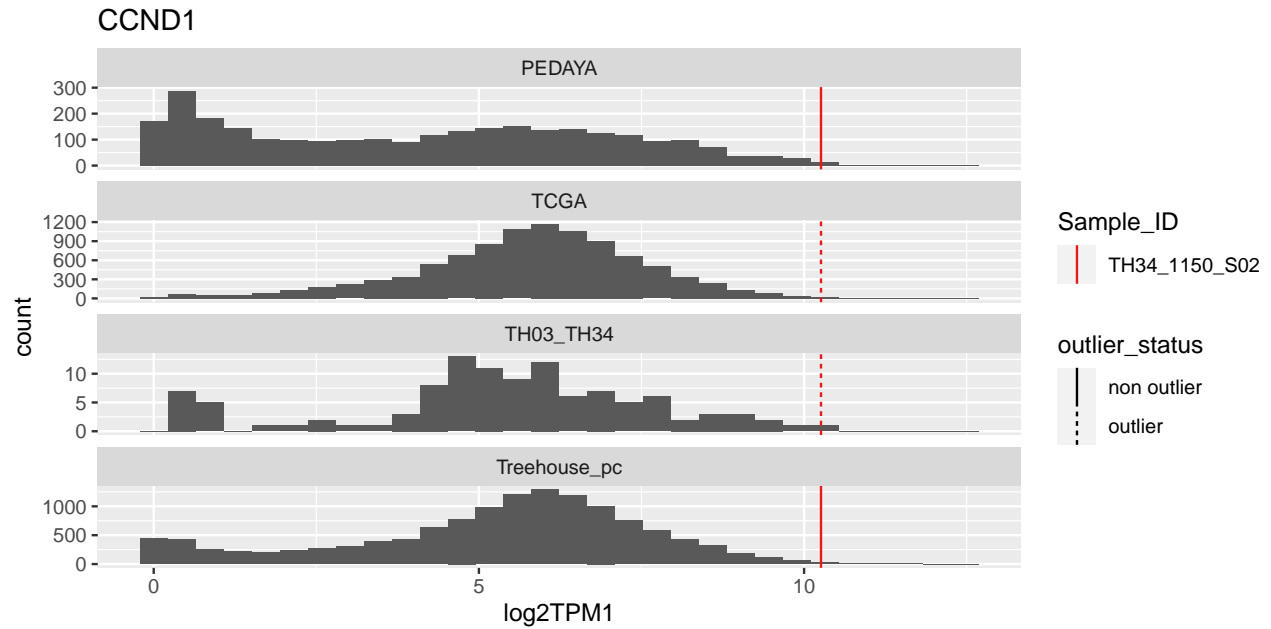
## [[36]]



Sample_ID	gene	PEDAYA	TCGA	TH03_TH34	Treehouse_pc	log2TPM1
TH34_1381_S01	VEGFA	FALSE	TRUE	TRUE	TRUE	11.031
TH34_1445_S02	VEGFA	FALSE	TRUE	FALSE	FALSE	10.196
TH34_1445_S02	VEGFA	FALSE	TRUE	FALSE	FALSE	NA
TH34_2292_S01	VEGFA	FALSE	TRUE	TRUE	TRUE	10.571
TH34_2292_S01	VEGFA	FALSE	TRUE	TRUE	TRUE	NA

##

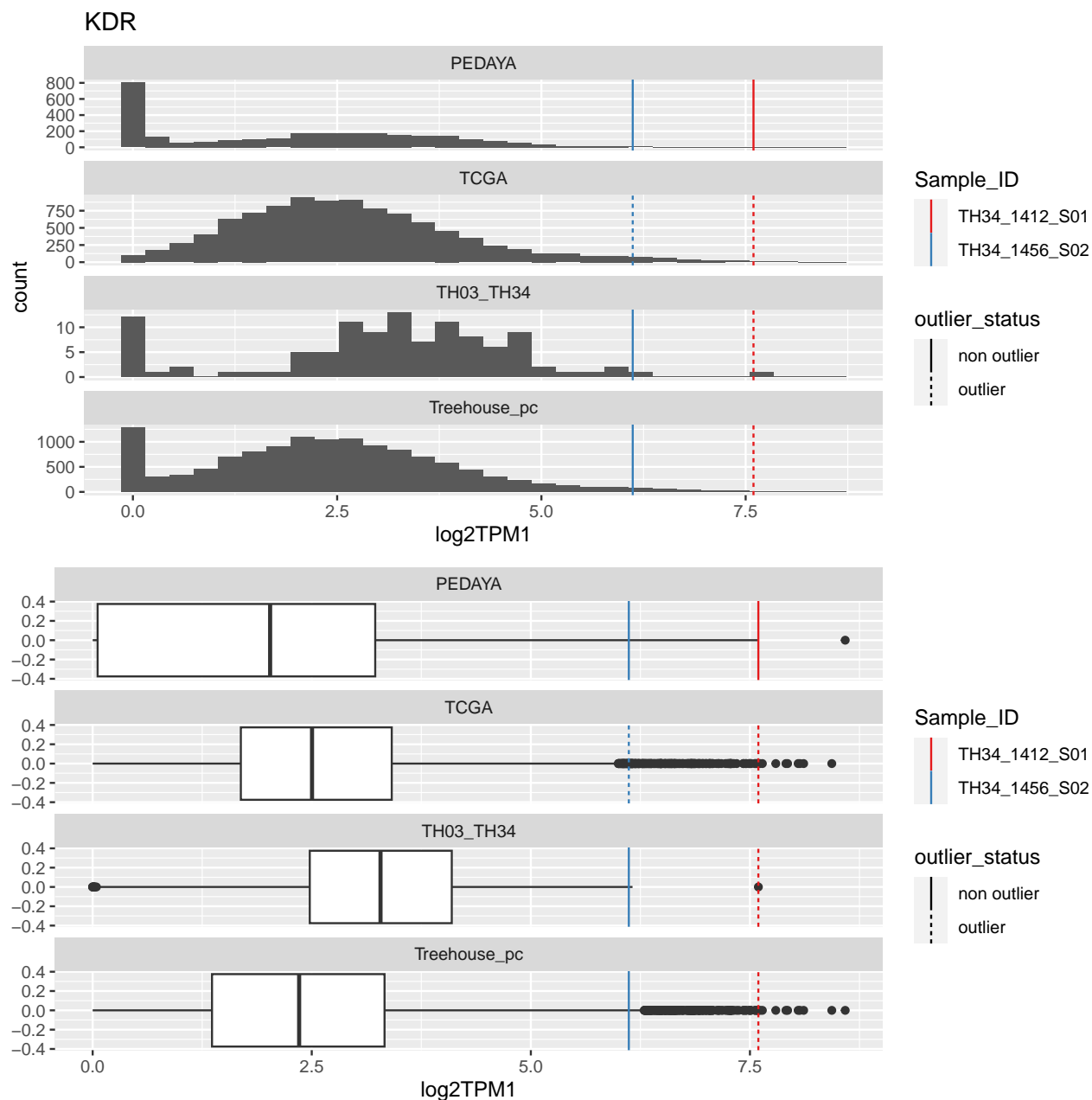
## [[37]]



Sample_ID	gene	PEDAYA	TCGA	TH03_TH34	Treehouse_pc	log2TPM1
TH34_1150_S02	CCND1	FALSE	TRUE	TRUE	FALSE	10.26



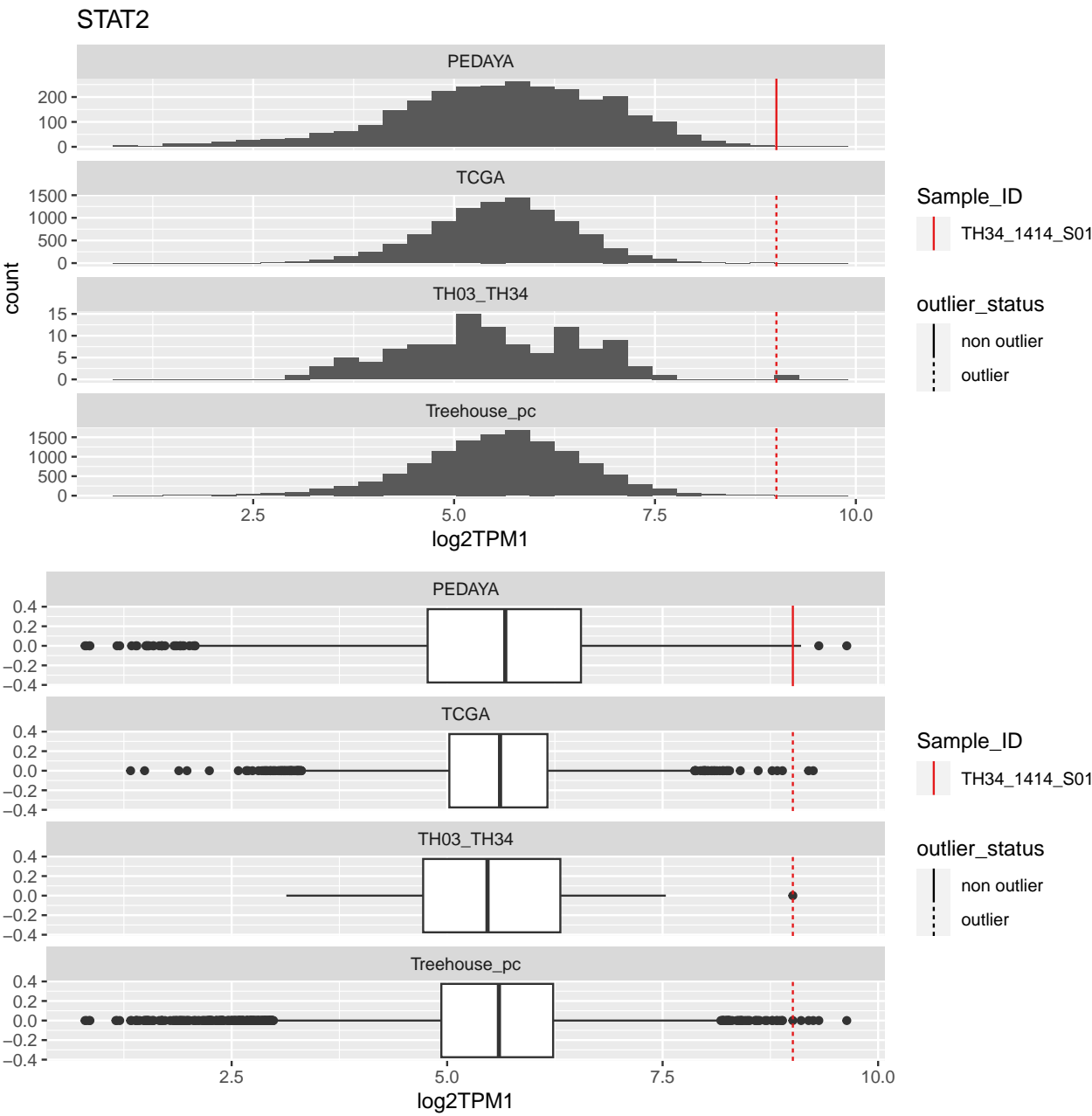
```
##  
## [[38]]
```



Sample_ID	gene	PEDAYA	TCGA	TH03_TH34	Treehouse_pc	log2TPM1
TH34_1412_S01	KDR	FALSE	TRUE	TRUE	TRUE	7.596
TH34_1456_S02	KDR	FALSE	TRUE	FALSE	FALSE	6.118
TH34_1456_S02	KDR	FALSE	TRUE	FALSE	FALSE	NA

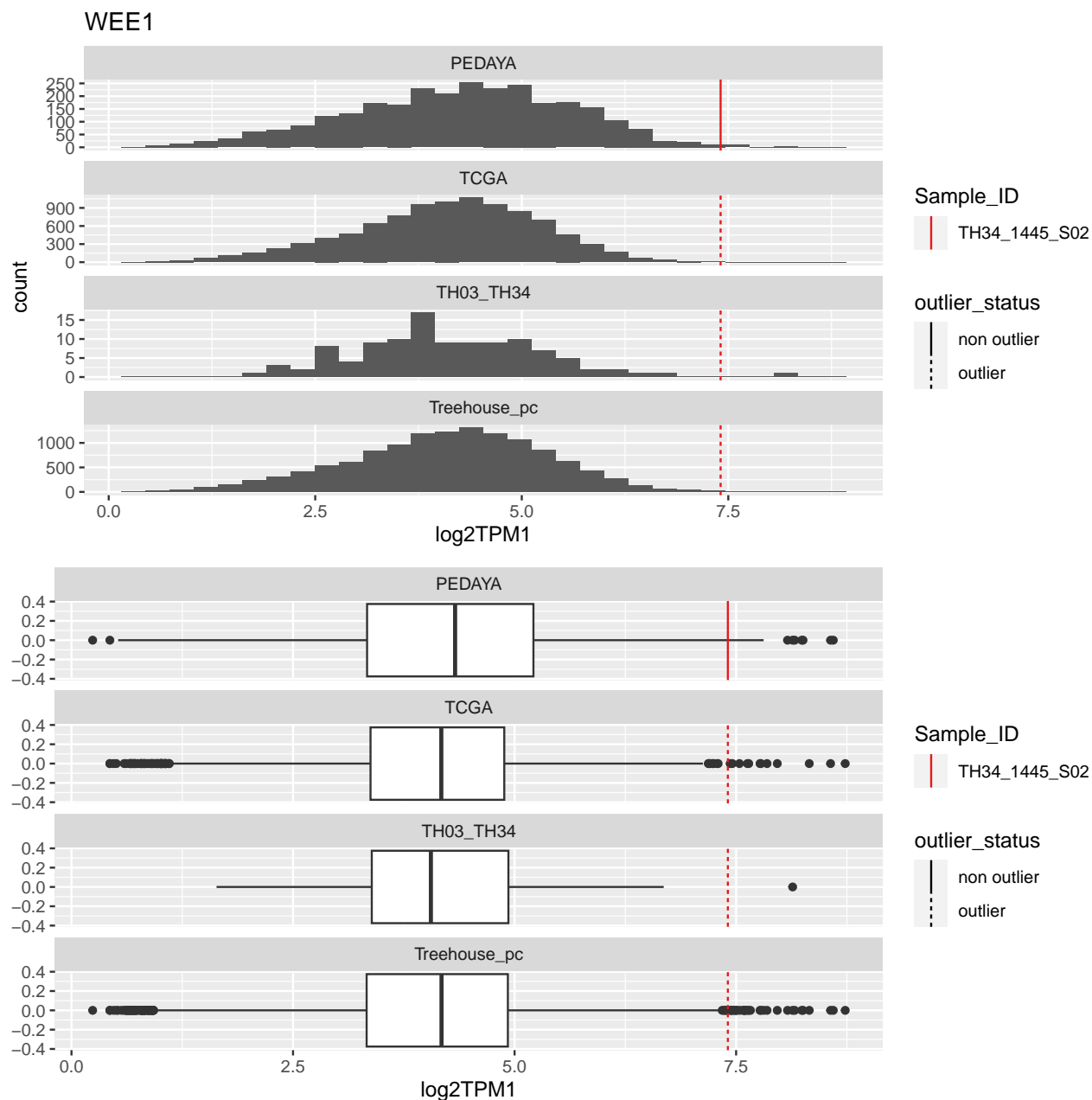
##

## [[39]]



Sample_ID	gene	PEDAYA	TCGA	TH03_TH34	Treehouse_pc	log2TPM1
TH34_1414_S01	STAT2	FALSE	TRUE	TRUE	TRUE	9.011

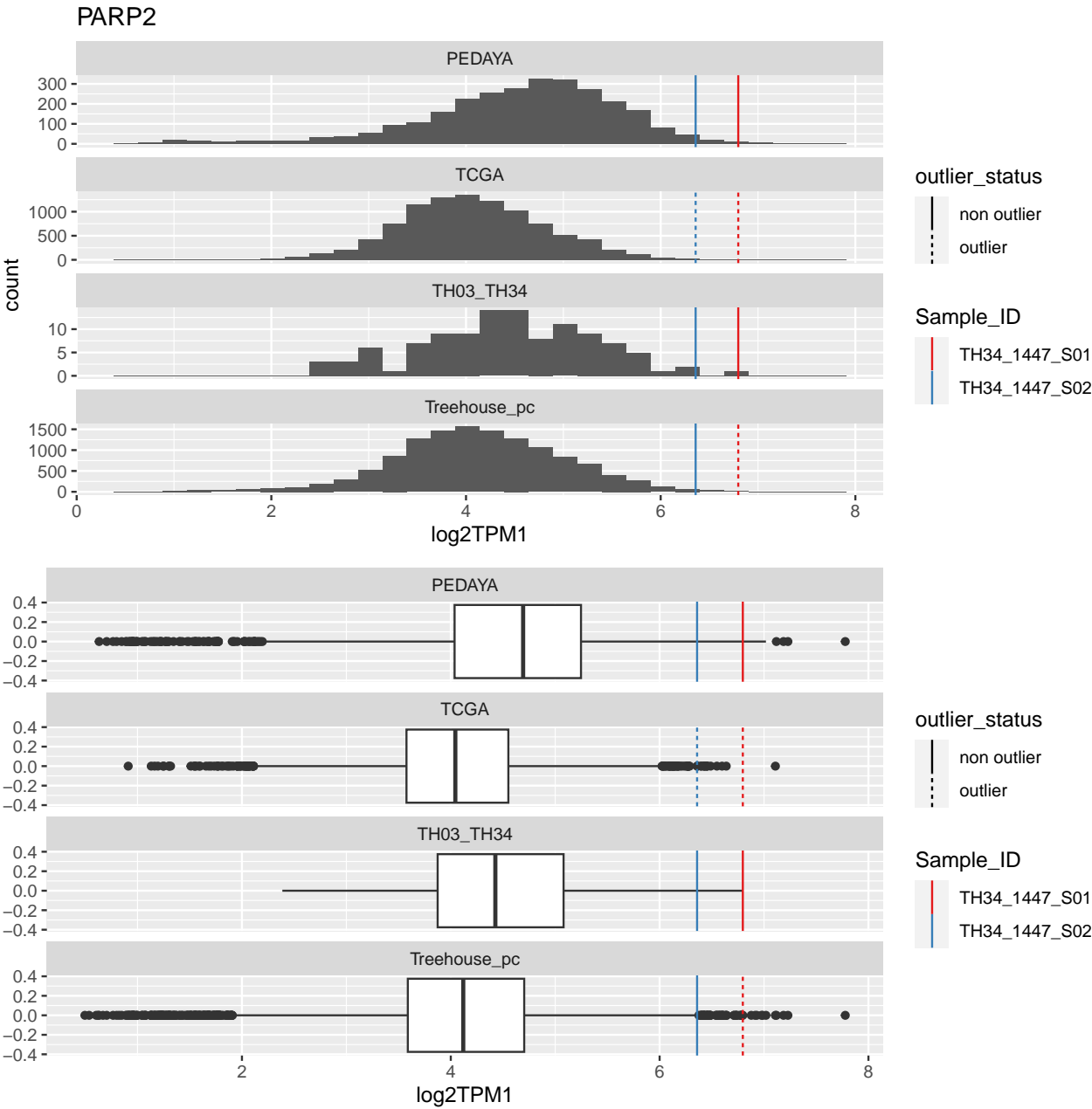
```
##  
## [[40]]
```



Sample_ID	gene	PEDAYA	TCGA	TH03_TH34	Treehouse_pc	log2TPM1
TH34_1445_S02	WEE1	FALSE	TRUE	TRUE	TRUE	7.407
TH34_1445_S02	WEE1	FALSE	TRUE	TRUE	TRUE	NA

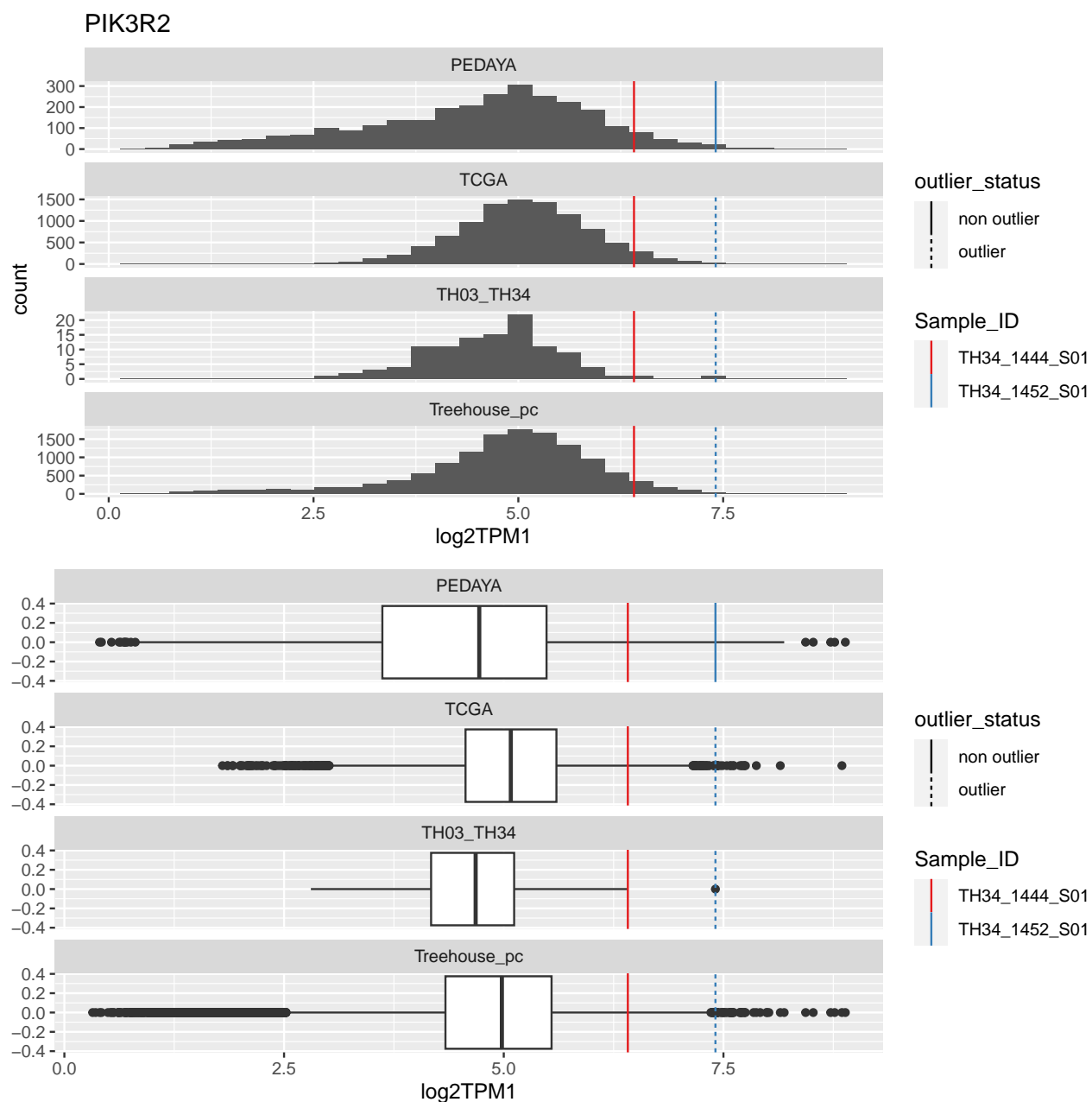
##

## [[41]]



Sample_ID	gene	PEDAYA	TCGA	TH03_TH34	Treehouse_pc	log2TPM1
TH34_1447_S01	PARP2	FALSE	TRUE	FALSE	TRUE	6.797
TH34_1447_S02	PARP2	FALSE	TRUE	FALSE	FALSE	6.359

```
##  
## [[42]]
```

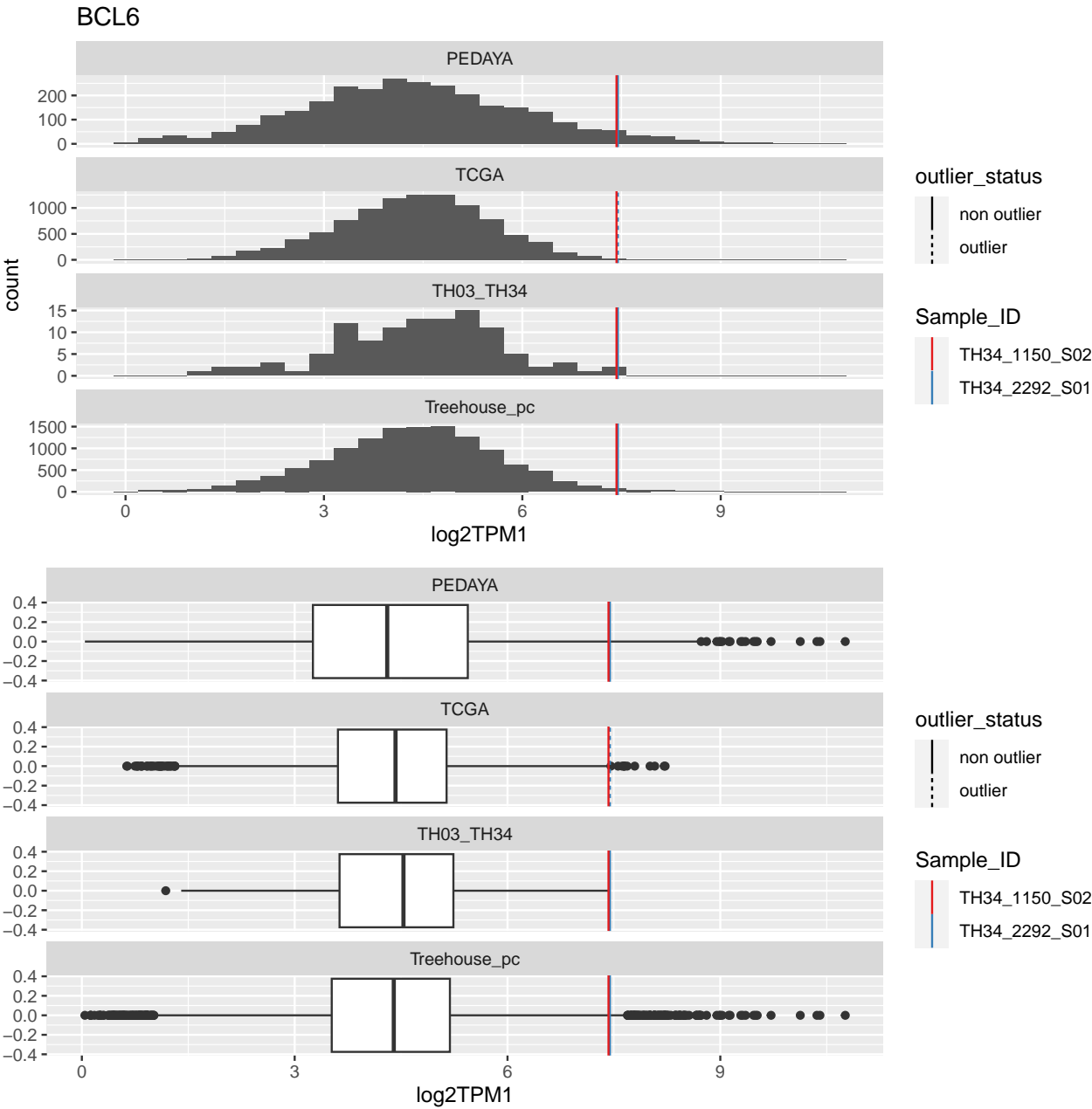


Sample_ID	gene	PEDAYA	TCGA	TH03_TH34	Treehouse_pc	log2TPM1
TH34_1452_S01	PIK3R2	FALSE	TRUE	TRUE	TRUE	7.409
TH34_1444_S01	PIK3R2	FALSE	FALSE	FALSE	FALSE	6.412

##

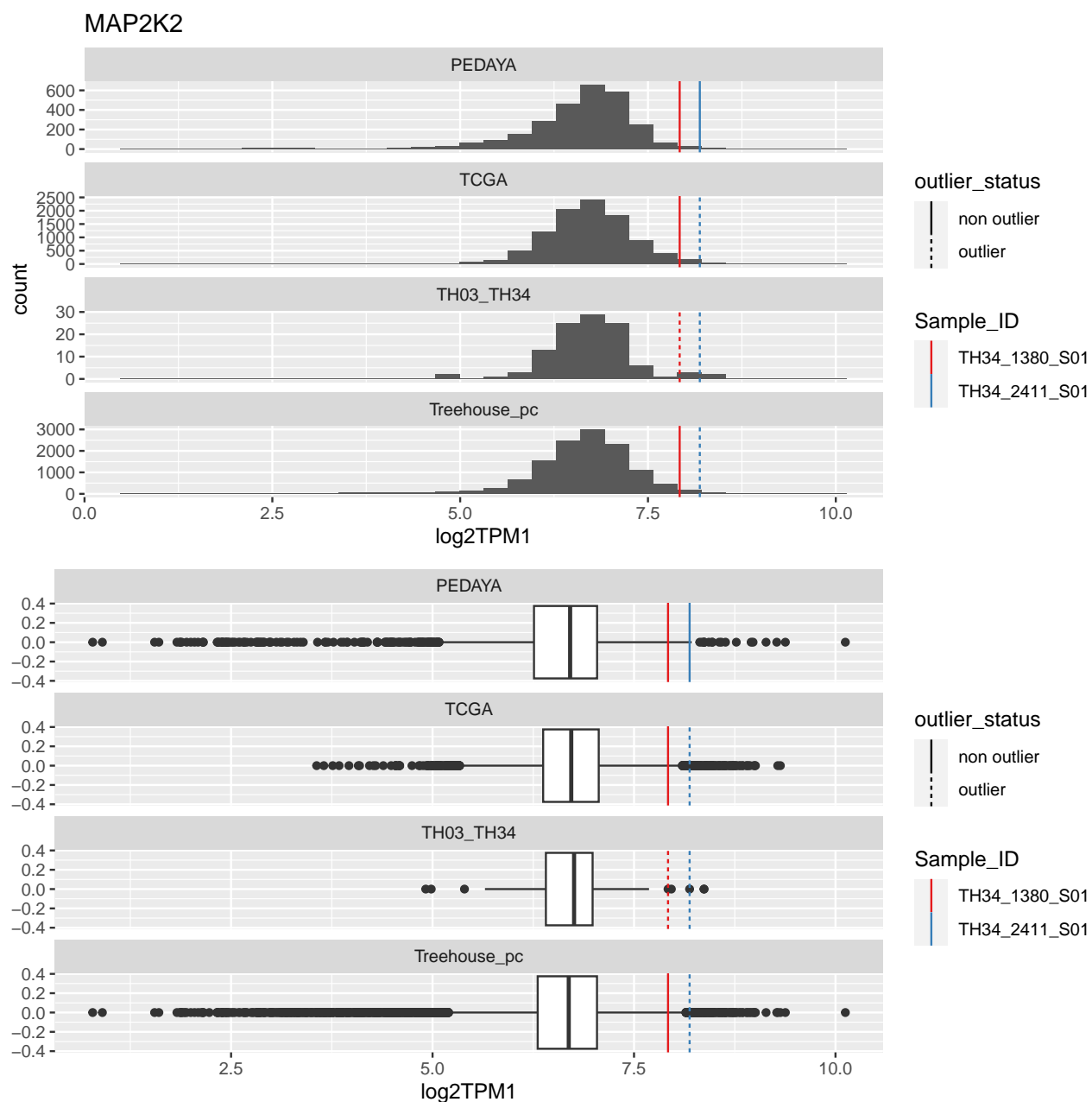


## [[43]]



Sample_ID	gene	PEDAYA	TCGA	TH03_TH34	Treehouse_pc	log2TPM1
TH34_2292_S01	BCL6	FALSE	TRUE	FALSE	FALSE	7.444
TH34_2292_S01	BCL6	FALSE	TRUE	FALSE	FALSE	NA
TH34_1150_S02	BCL6	FALSE	FALSE	FALSE	FALSE	7.424

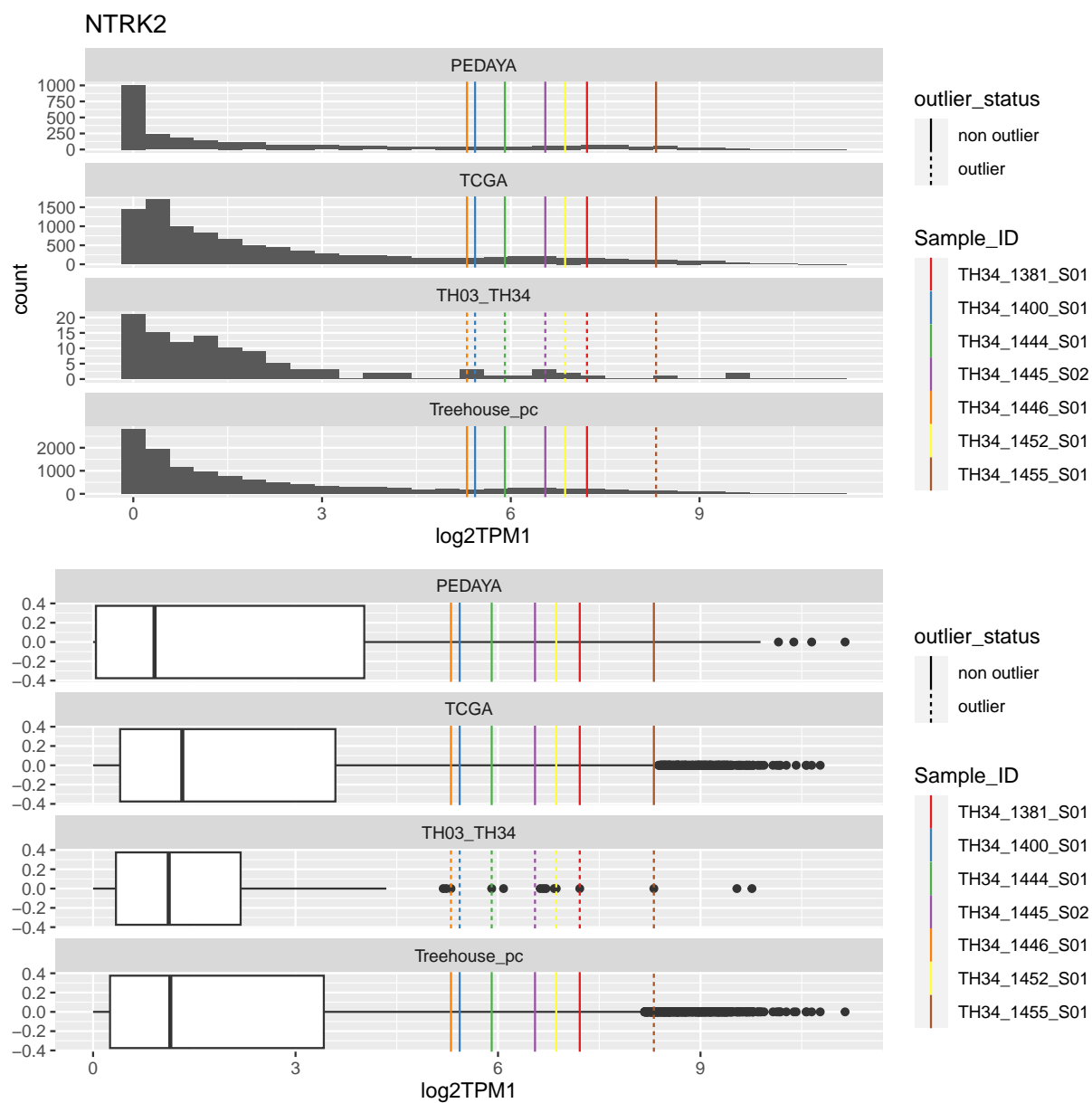
```
##  
## [[44]]
```



Sample_ID	gene	PEDAYA	TCGA	TH03_TH34	Treehouse_pc	log2TPM1
TH34_2411_S01	MAP2K2	FALSE	TRUE	TRUE	TRUE	8.190
TH34_1380_S01	MAP2K2	FALSE	FALSE	TRUE	FALSE	7.922

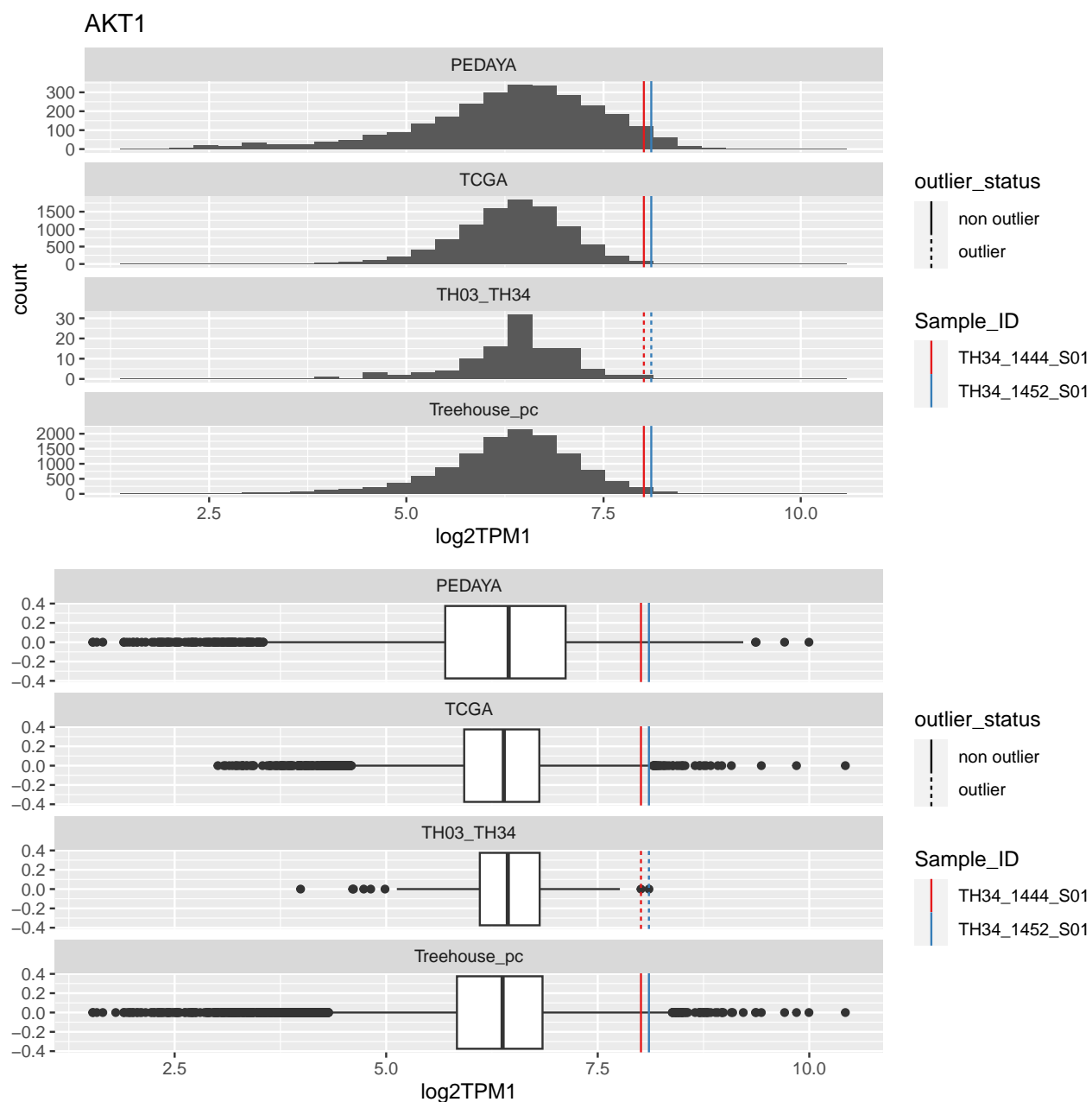
##

## [[45]]



Sample_ID	gene	PEDAYA	TCGA	TH03_TH34	Treehouse_pc	log2TPM1
TH34_1400_S01	NTRK2	FALSE	FALSE	TRUE	FALSE	5.432
TH34_1400_S01	NTRK2	FALSE	FALSE	TRUE	FALSE	NA
TH34_1444_S01	NTRK2	FALSE	FALSE	TRUE	FALSE	5.906
TH34_1445_S02	NTRK2	FALSE	FALSE	TRUE	FALSE	6.548
TH34_1445_S02	NTRK2	FALSE	FALSE	TRUE	FALSE	NA
TH34_1446_S01	NTRK2	FALSE	FALSE	TRUE	FALSE	5.303
TH34_1452_S01	NTRK2	FALSE	FALSE	TRUE	FALSE	6.863
TH34_1455_S01	NTRK2	FALSE	FALSE	TRUE	TRUE	8.309
TH34_1381_S01	NTRK2	FALSE	FALSE	TRUE	FALSE	7.211

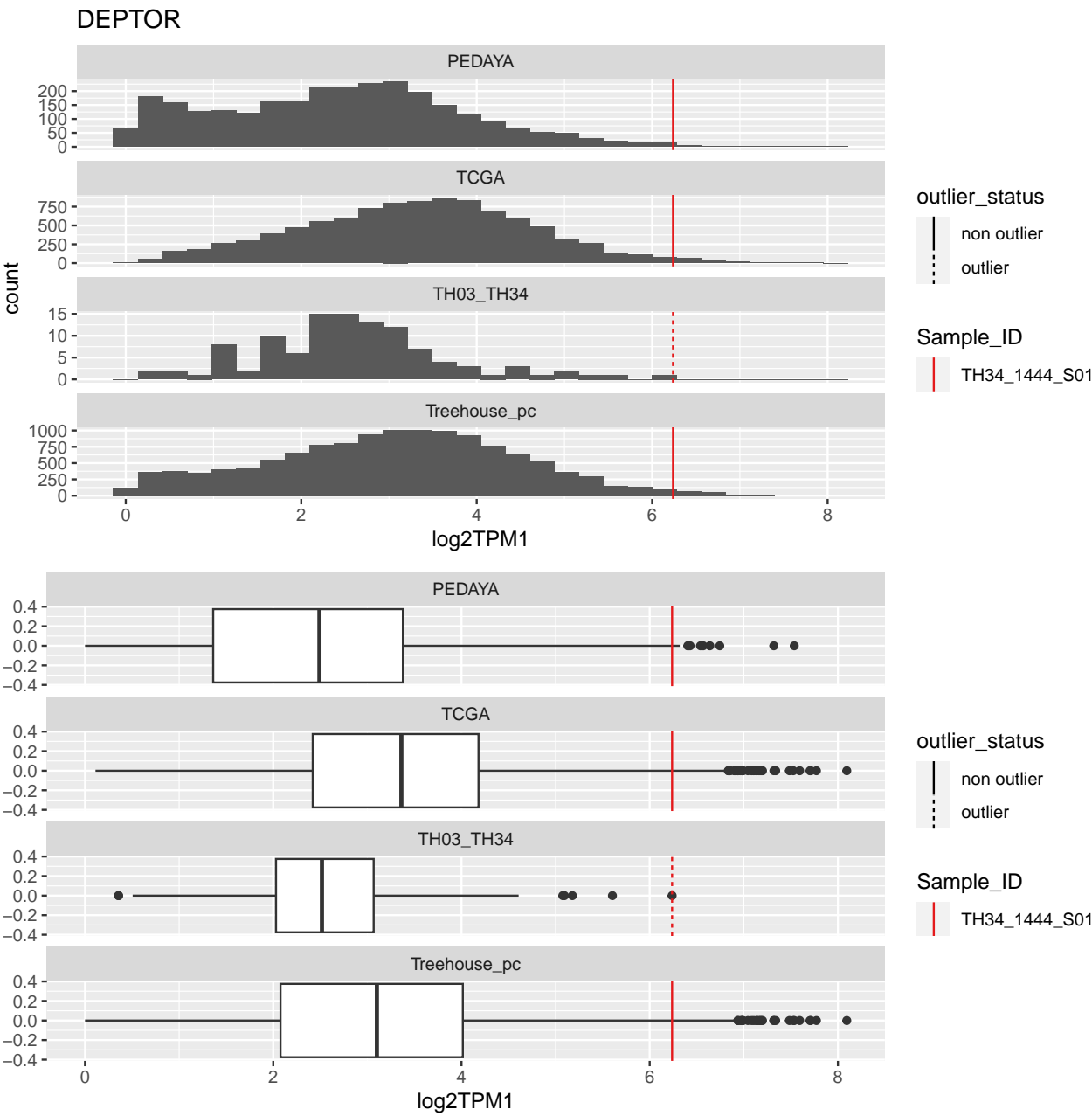
```
##  
## [[46]]
```



Sample_ID	gene	PEDAYA	TCGA	TH03_TH34	Treehouse_pc	log2TPM1
TH34_1444_S01	AKT1	FALSE	FALSE	TRUE	FALSE	8.011
TH34_1452_S01	AKT1	FALSE	FALSE	TRUE	FALSE	8.105

##

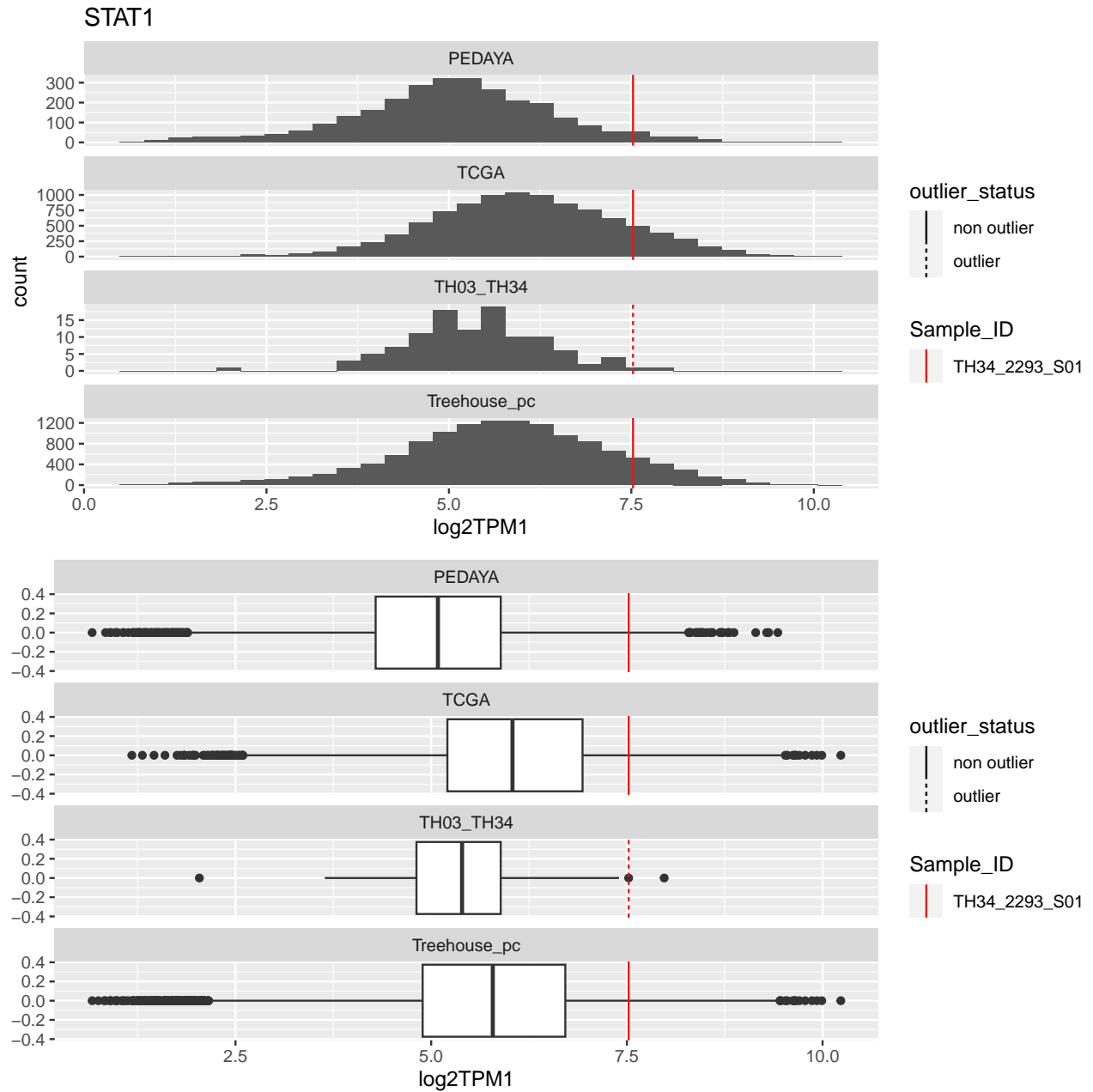
## [[47]]



Sample_ID	gene	PEDAYA	TCGA	TH03_TH34	Treehouse_pc	log2TPM1
TH34_1444_S01	DEPTOR	FALSE	FALSE	TRUE	FALSE	6.238



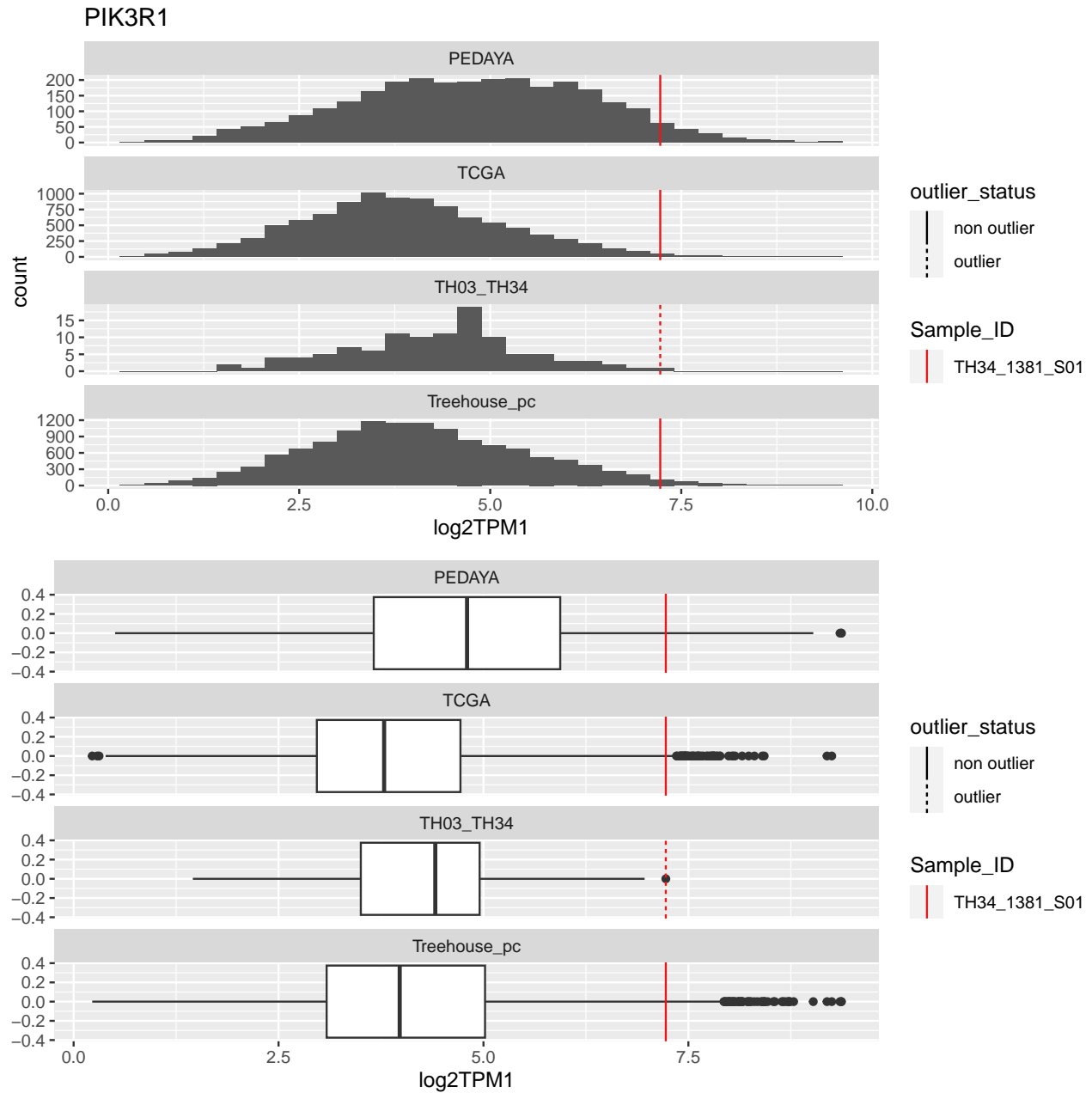
```
##
## [[48]]
```



Sample_ID	gene	PEDAYA	TCGA	TH03_TH34	Treehouse_pc	log2TPM1
TH34_2293_S01	STAT1	FALSE	FALSE	TRUE	FALSE	7.523

##

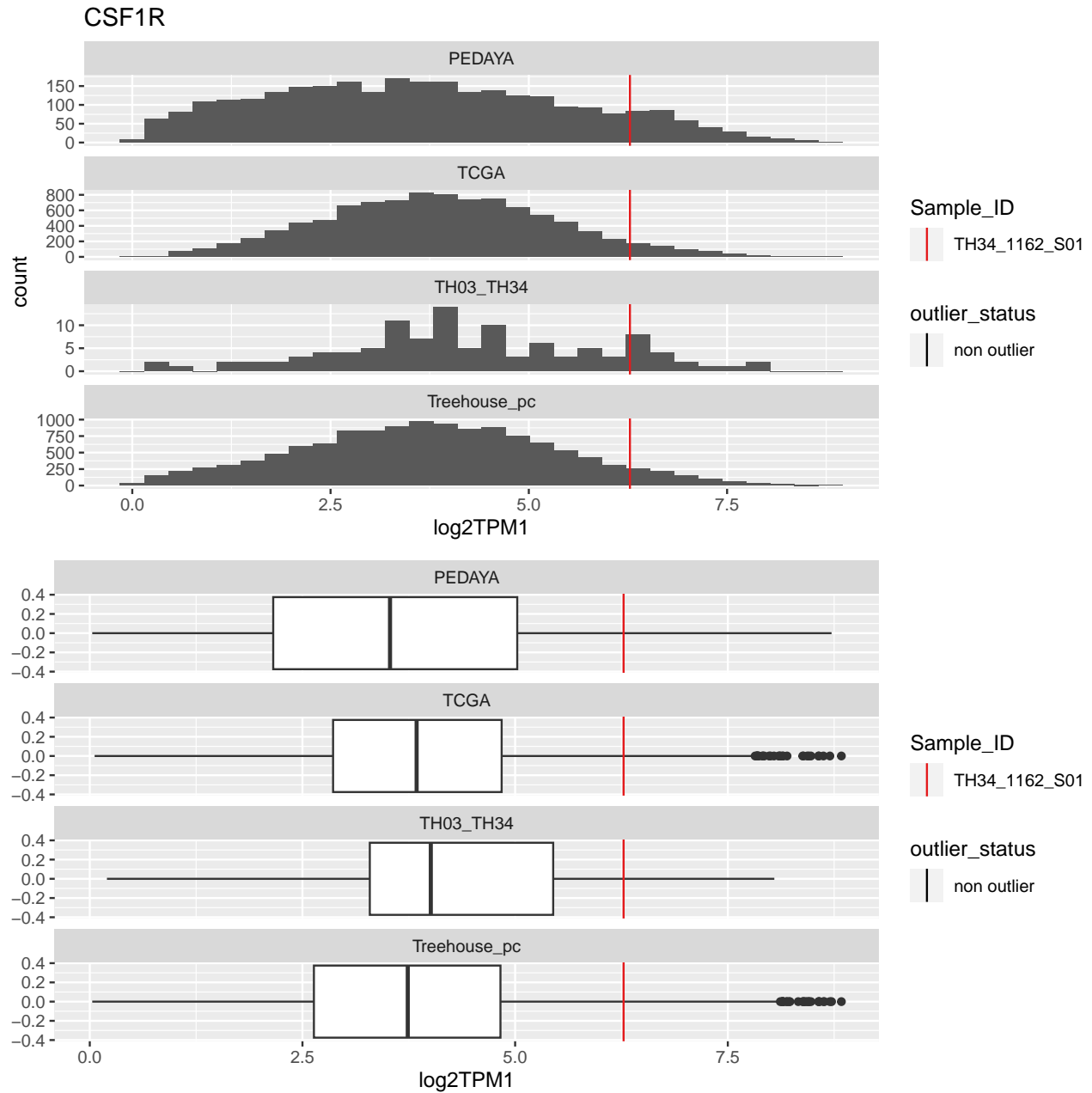
## [[49]]



Sample_ID	gene	PEDAYA	TCGA	TH03_TH34	Treehouse_pc	log2TPM1
TH34_1381_S01	PIK3R1	FALSE	FALSE	TRUE	FALSE	7.225

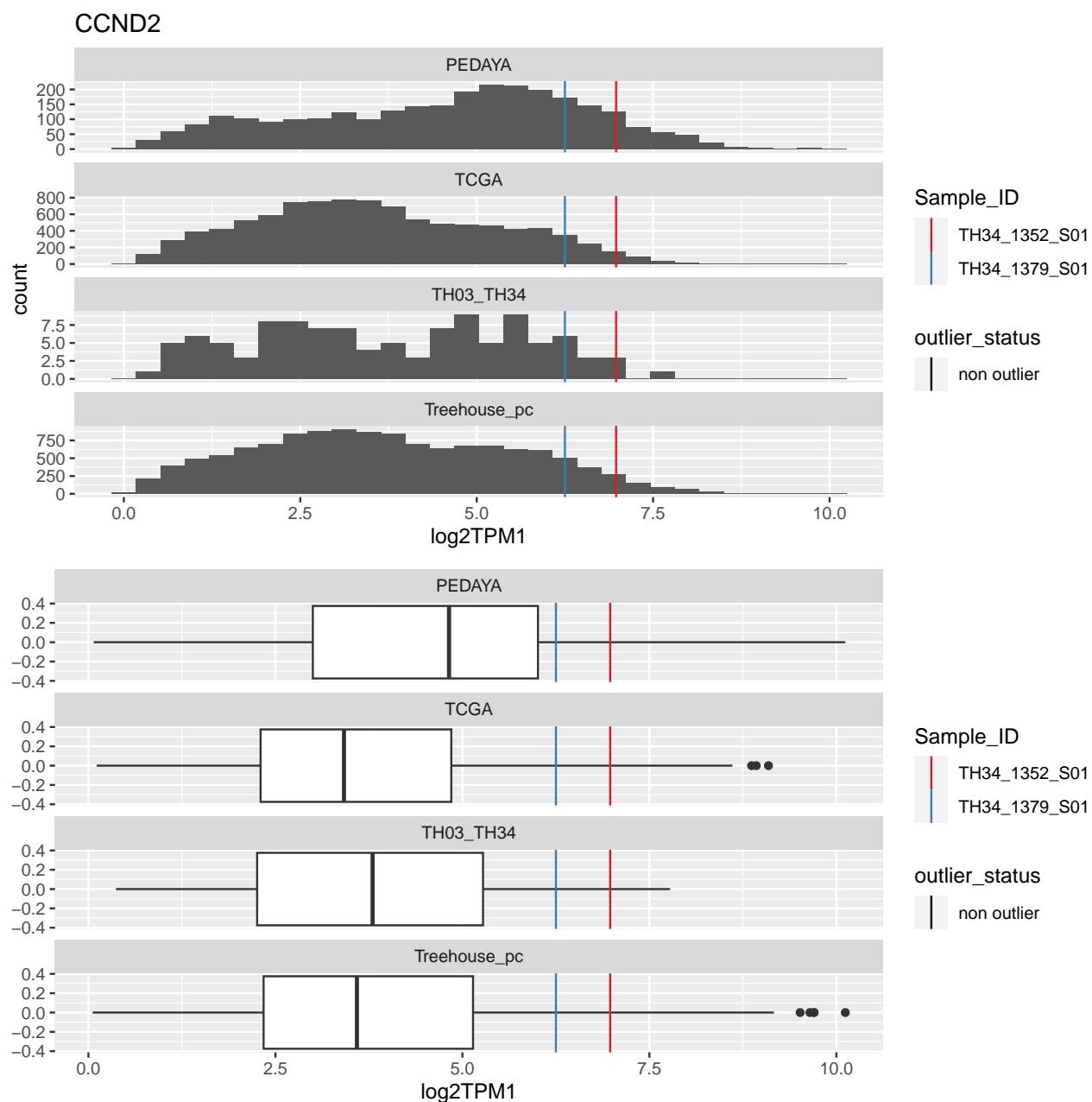
##

## [[50]]



Sample_ID	gene	PEDAYA	TCGA	TH03_TH34	Treehouse_pc	log2TPM1
TH34_1162_S01	CSF1R	FALSE	FALSE	FALSE	FALSE	6.275

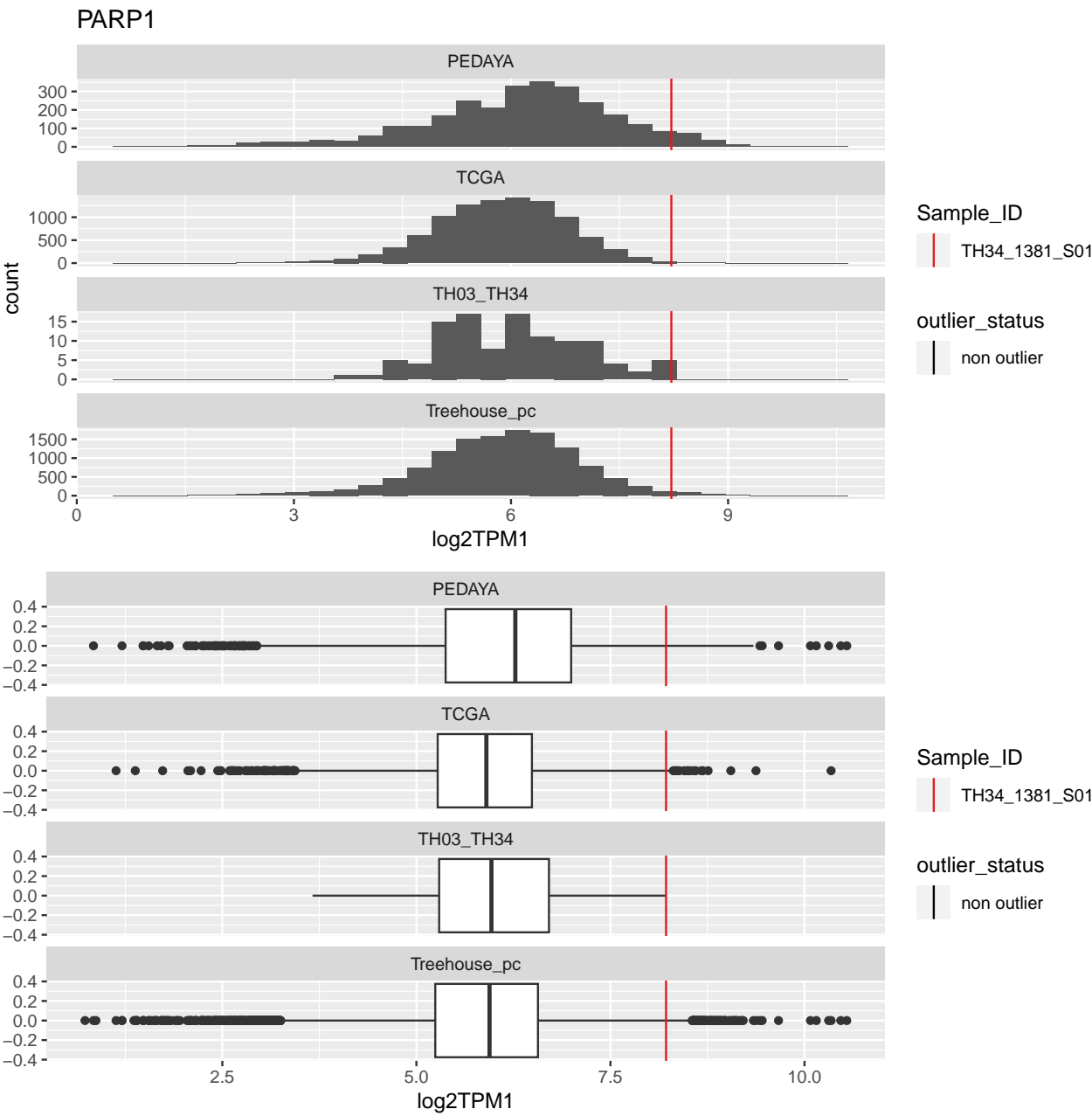
```
##  
## [[51]]
```



Sample_ID	gene	PEDAYA	TCGA	TH03_TH34	Treehouse_pc	log2TPM1
TH34_1352_S01	CCND2	FALSE	FALSE	FALSE	FALSE	6.975
TH34_1379_S01	CCND2	FALSE	FALSE	FALSE	FALSE	6.250

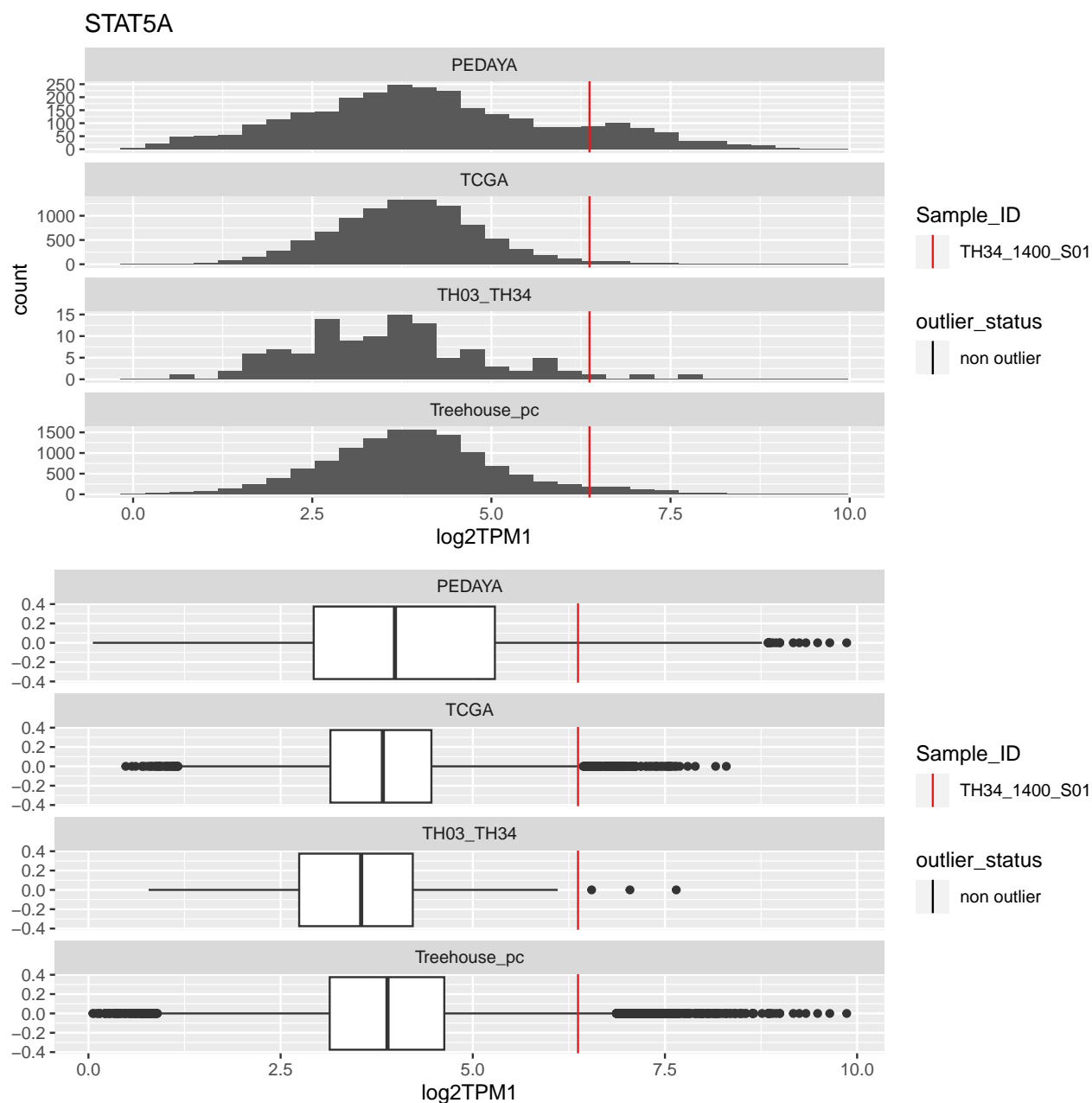
##

## [[52]]



Sample_ID	gene	PEDAYA	TCGA	TH03_TH34	Treehouse_pc	log2TPM1
TH34_1381_S01	PARP1	FALSE	FALSE	FALSE	FALSE	8.216

```
##  
## [[53]]
```

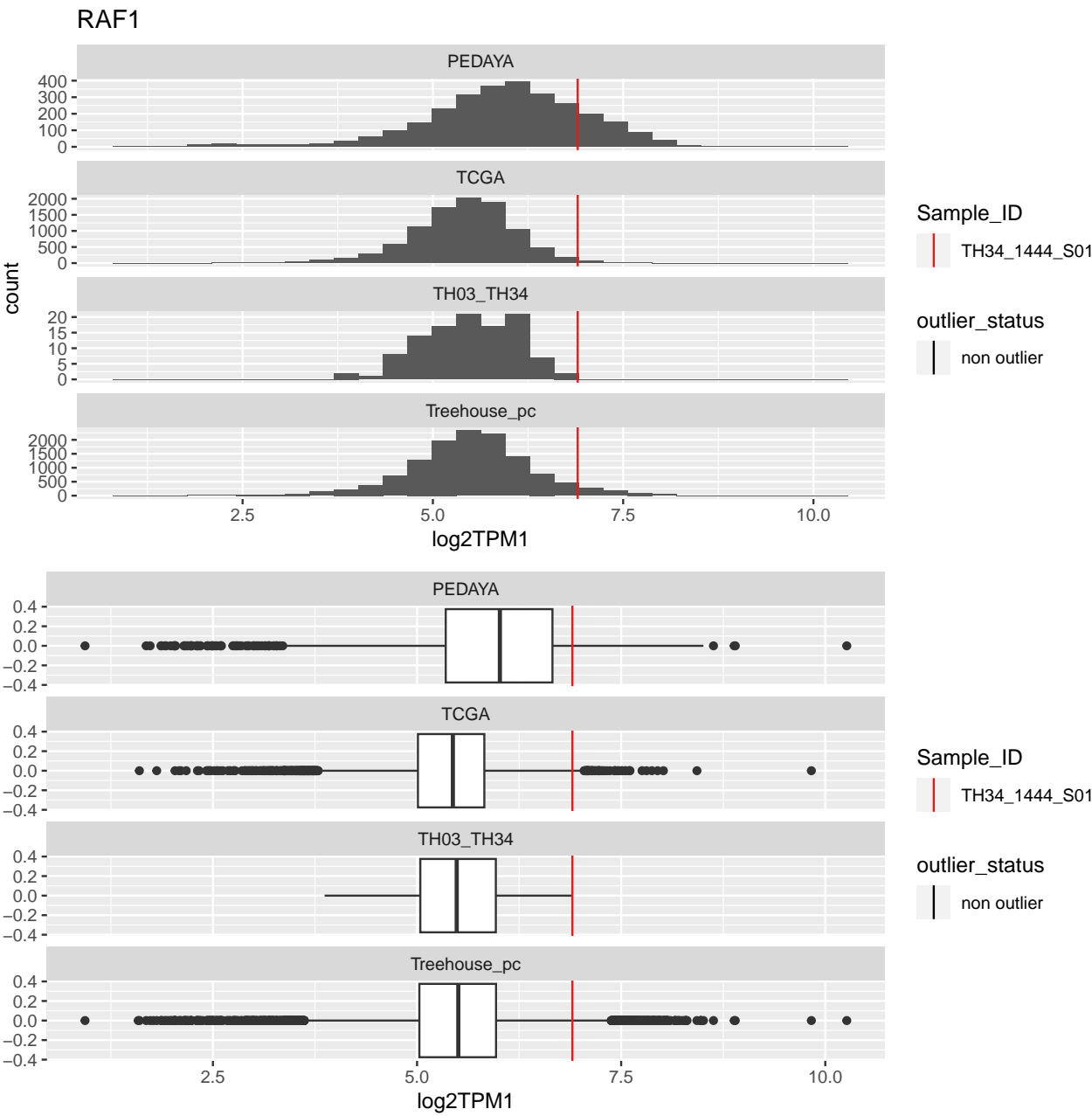


Sample_ID	gene	PEDAYA	TCGA	TH03_TH34	Treehouse_pc	log2TPM1
TH34_1400_S01	STAT5A	FALSE	FALSE	FALSE	FALSE	6.369
TH34_1400_S01	STAT5A	FALSE	FALSE	FALSE	FALSE	NA

##



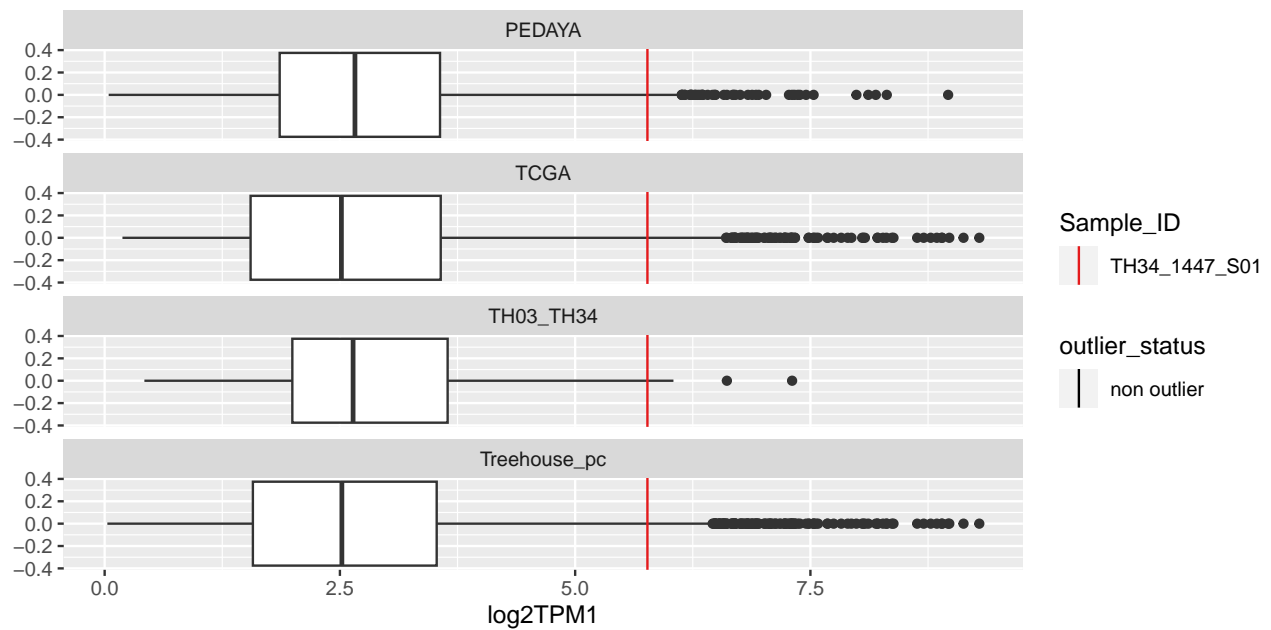
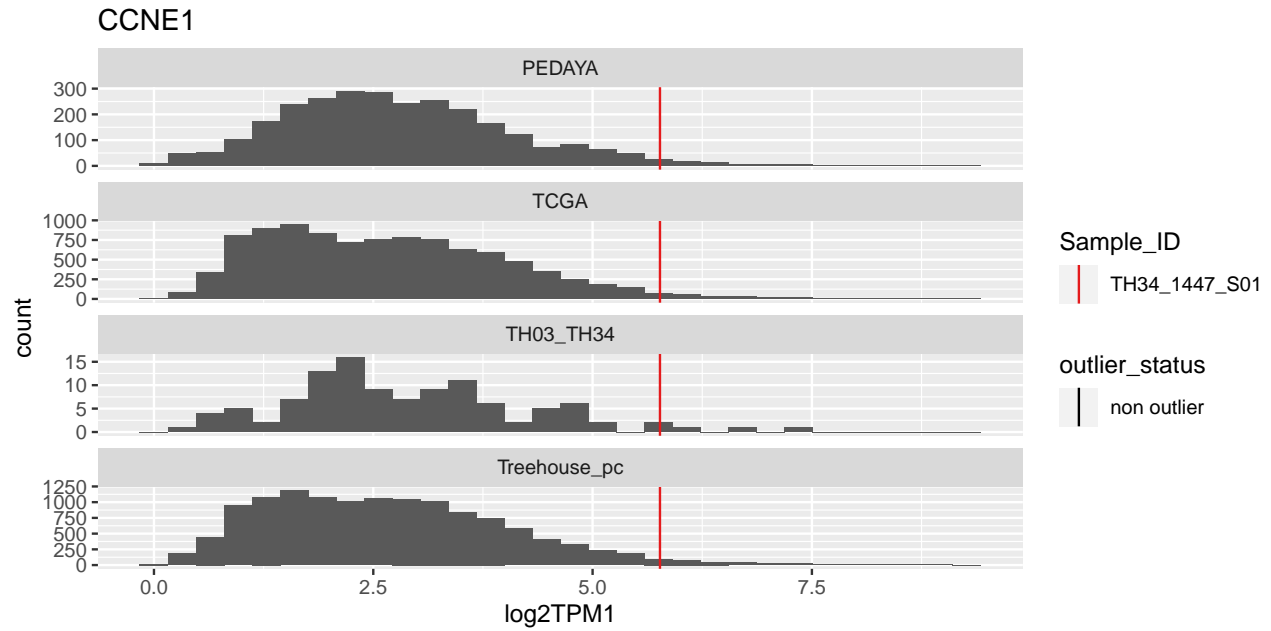
## [[54]]



Sample_ID	gene	PEDAYA	TCGA	TH03_TH34	Treehouse_pc	log2TPM1
TH34_1444_S01	RAF1	FALSE	FALSE	FALSE	FALSE	6.903

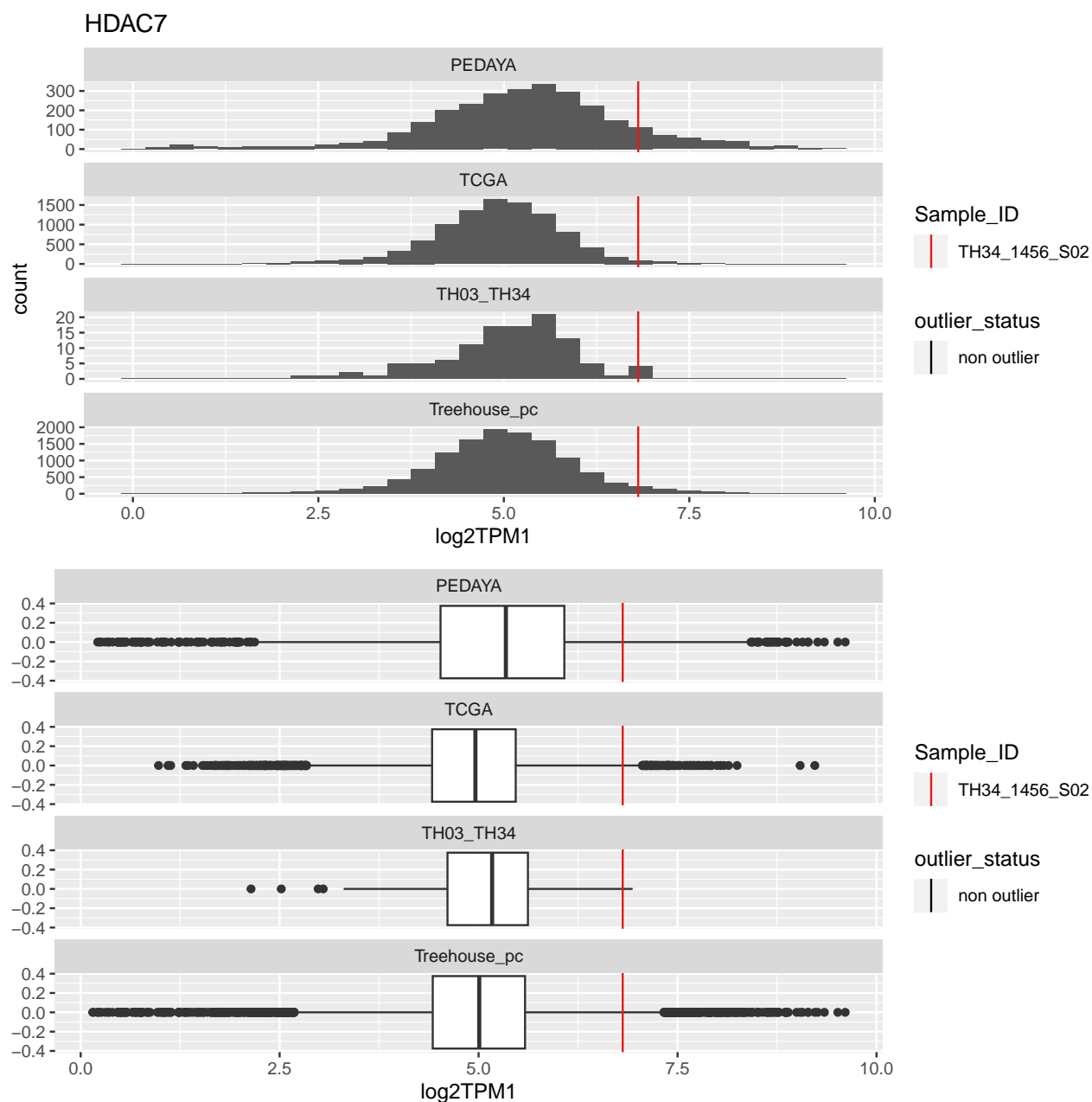
##

## [[55]]



Sample_ID	gene	PEDAYA	TCGA	TH03_TH34	Treehouse_pc	log2TPM1
TH34_1447_S01	CCNE1	FALSE	FALSE	FALSE	FALSE	5.768

```
##  
## [[56]]
```



Sample_ID	gene	PEDAYA	TCGA	TH03_TH34	Treehouse_pc	log2TPM1
TH34_1456_S02	HDAC7	FALSE	FALSE	FALSE	FALSE	6.811
TH34_1456_S02	HDAC7	FALSE	FALSE	FALSE	FALSE	NA