review-expression-distributions-in-different-comparison cohorts 2023.11.16 13.58.16

hbeale

November 17, 2023

Contents


```
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
                           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_
                           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.samp
                           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")) %>%
  na.omit() # excludes samples not in compendium

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: 390 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 380 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(sort(outlier_genes_detected), function(this_gene){
  # this_gene <- "BCL6"
  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])

 cohort_thresholds <-  one_gene_expr_per_cohort %>%
  group_by(cohort) %>%
  summarize(q25 = quantile(log2TPM1, 0.25),
            median = median(log2TPM1),
            q75 = quantile(log2TPM1, 0.75),
            IQR = q75-q25,
            up_outlier_threshold = q75 + (1.5*IQR)) %>%
  pivot_longer(-cohort) %>%
  mutate(value = round(value, 2)) %>%
  pivot_wider()


   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") +
```

```r
  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

  t2 <- tableGrob(cohort_thresholds, theme=ttheme_minimal(),
                  rows=NULL)  # transform into a tableGrob

  plot_grid(p1, p2, t1, t2,
                     ncol = 1)
})
```

```
## 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`.
## 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"
```

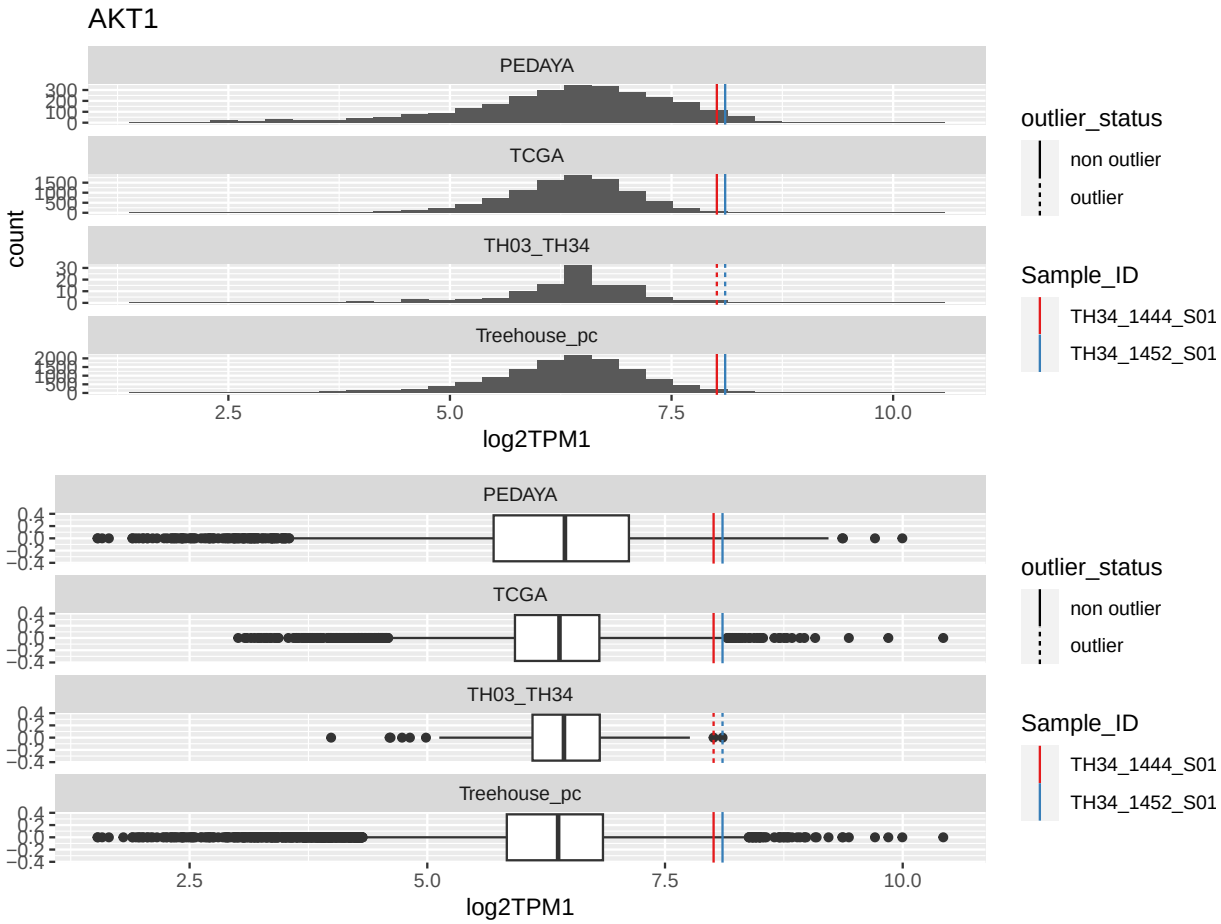
```
## `stat_bin()` using `bins = 30`. Pick better value with `binwidth`.
## Joining, by = "Sample_ID"
## `stat_bin()` using `bins = 30`. Pick better value with `binwidth`.
## Joining, by = "Sample_ID"
## `stat_bin()` using `bins = 30`. Pick better value with `binwidth`.
## Joining, by = "Sample_ID"
## `stat_bin()` using `bins = 30`. Pick better value with `binwidth`.
## 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 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): n too large, allowed maximum for palette Set1 is 9
## Returning the palette you asked for with that many colors
```

```
## 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`.
## 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"
```

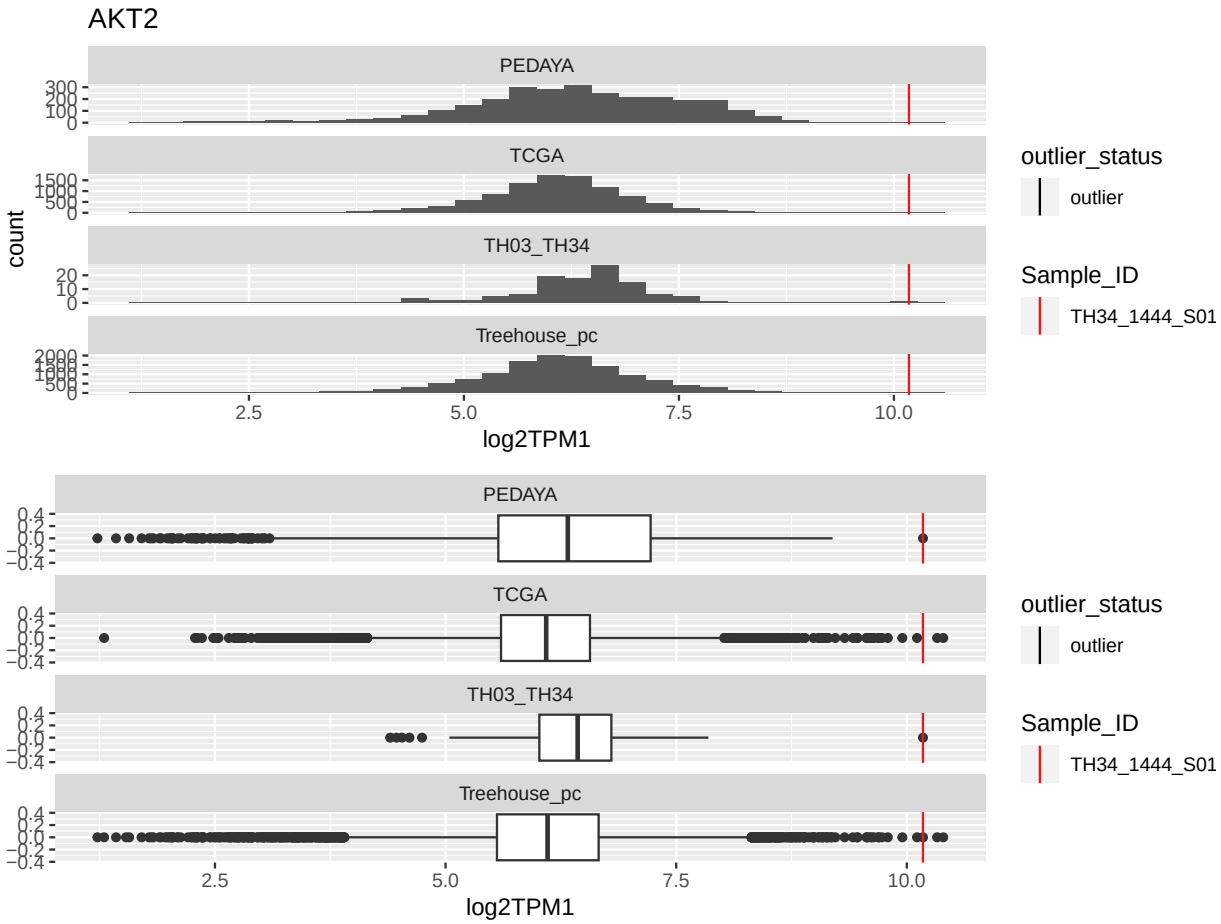
```
## `stat_bin()` using `bins = 30`. Pick better value with `binwidth`.
## Joining, by = "Sample_ID"
## `stat_bin()` using `bins = 30`. Pick better value with `binwidth`.
## Joining, by = "Sample_ID"
## `stat_bin()` using `bins = 30`. Pick better value with `binwidth`.
## Joining, by = "Sample_ID"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

## [[1]]
```



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

cohort	q25	median	q75	IQR	up_outlier_threshold
PEDAYA	5.70	6.45	7.12	1.42	9.25
TCGA	5.92	6.39	6.81	0.89	8.14
TH03_TH34	6.11	6.44	6.81	0.71	7.87
Treehouse_pc	5.84	6.38	6.85	1.01	8.36

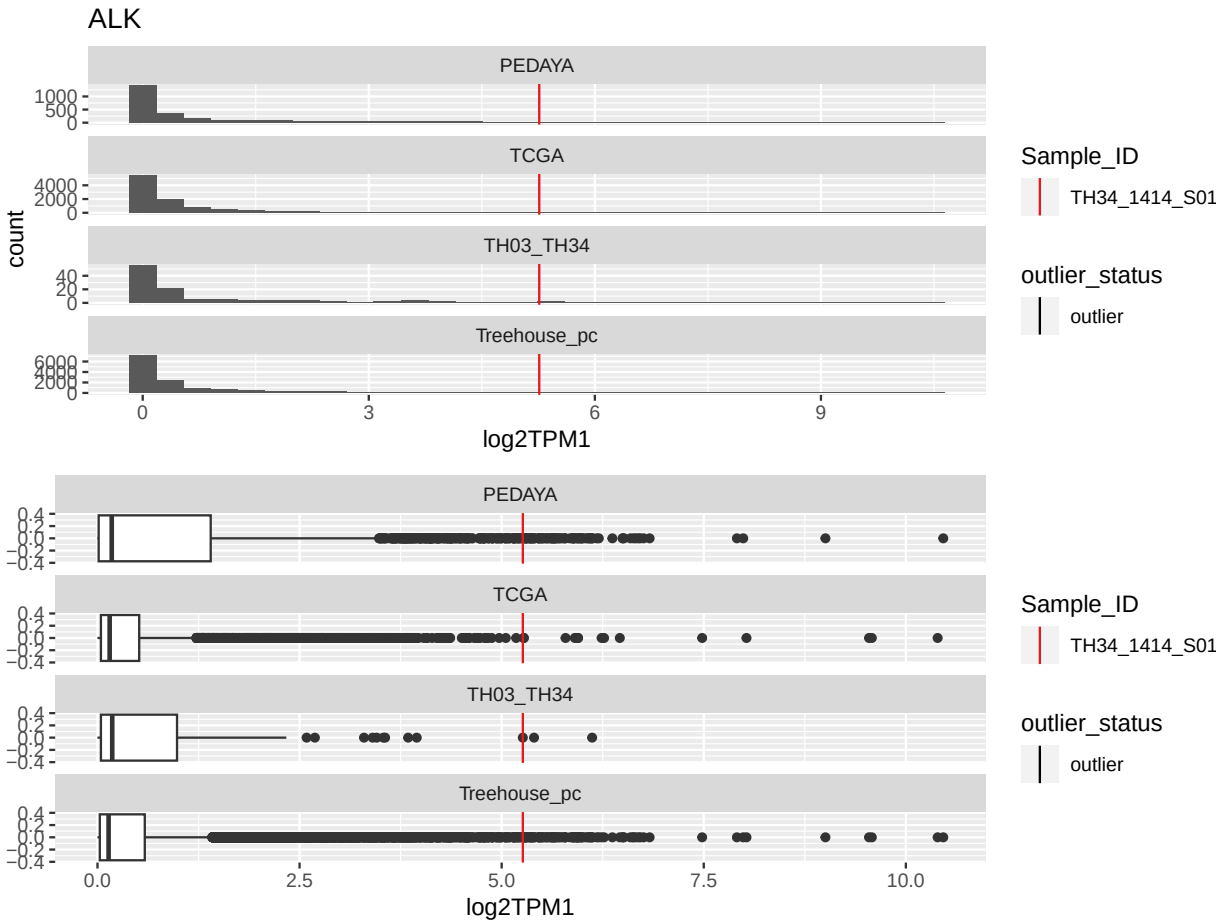
```
##
## [[2]]
```



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

cohort	q25	median	q75	IQR	up_outlier_threshold
PEDAYA	5.57	6.32	7.22	1.65	9.70
TCGA	5.60	6.09	6.57	0.96	8.01
TH03_TH34	6.02	6.43	6.80	0.78	7.97
Treehouse_pc	5.56	6.11	6.66	1.10	8.31

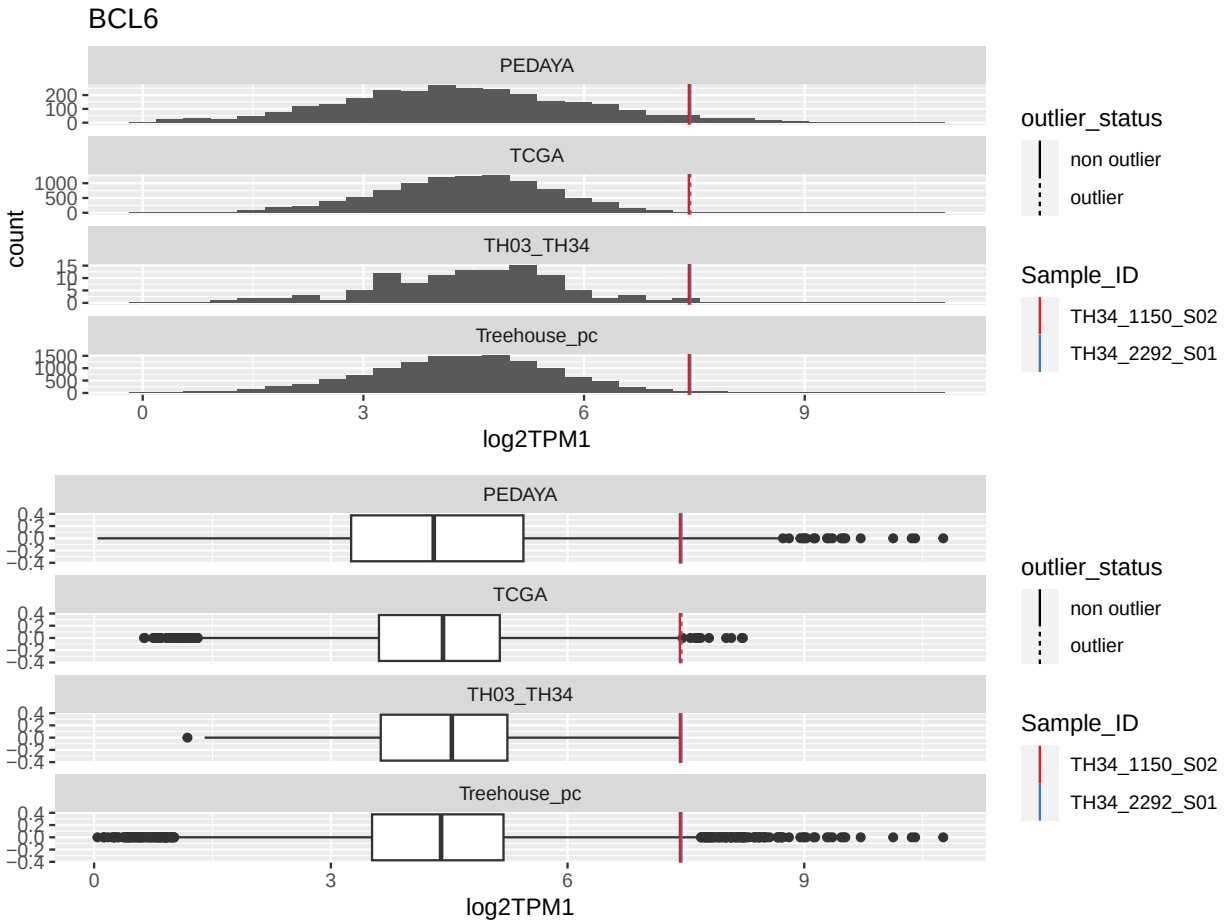
```
##
## [[3]]
```



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

cohort	q25	median	q75	IQR	up_outlier_threshold
PEDAYA	0.01	0.18	1.40	1.39	3.48
TCGA	0.04	0.15	0.52	0.47	1.23
TH03_TH34	0.04	0.18	0.98	0.94	2.40
Treehouse_pc	0.03	0.14	0.59	0.56	1.42

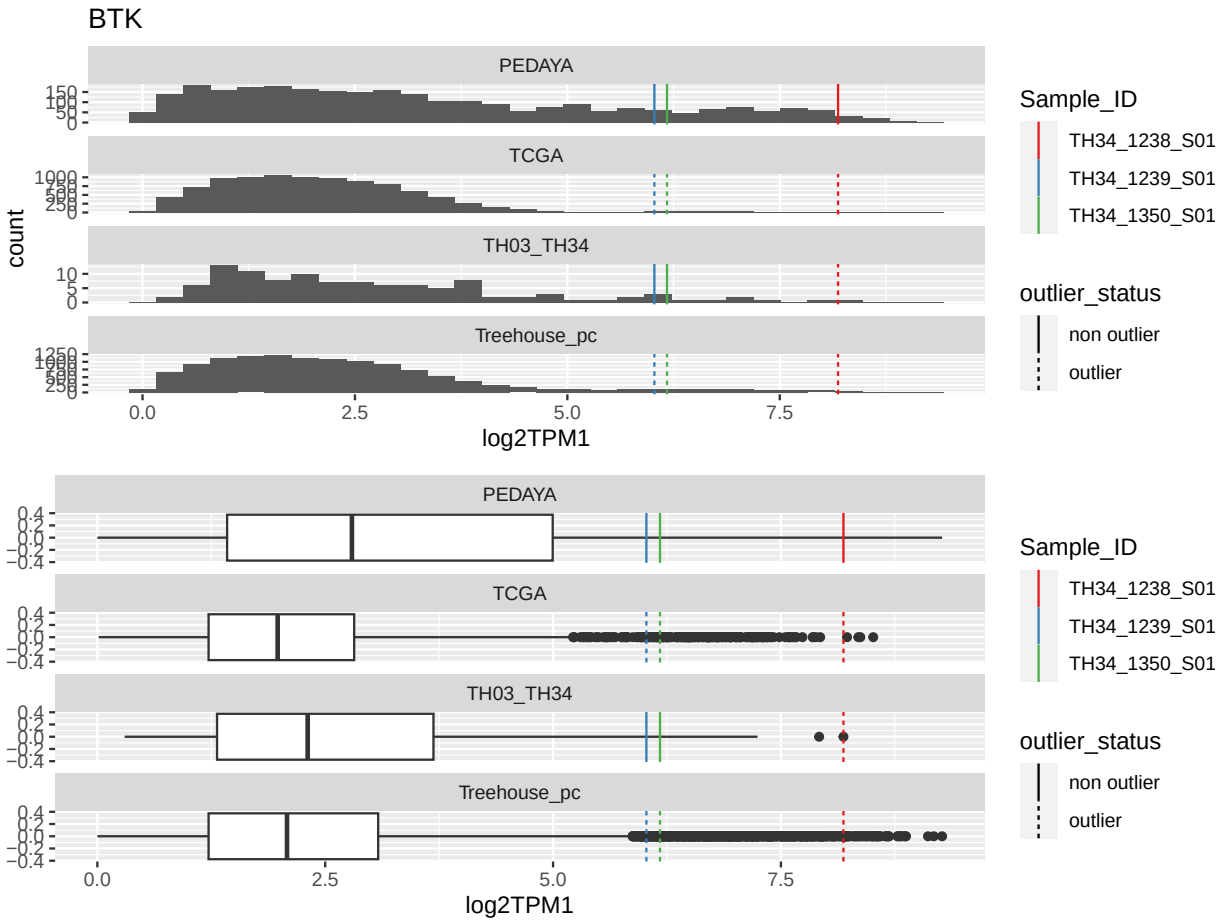
```
##
## [[4]]
```



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

cohort	q25	median	q75	IQR	up_outlier_threshold
PEDAYA	3.26	4.30	5.44	2.18	8.72
TCGA	3.61	4.42	5.14	1.53	7.44
TH03_TH34	3.63	4.53	5.24	1.61	7.65
Treehouse_pc	3.52	4.40	5.19	1.67	7.69

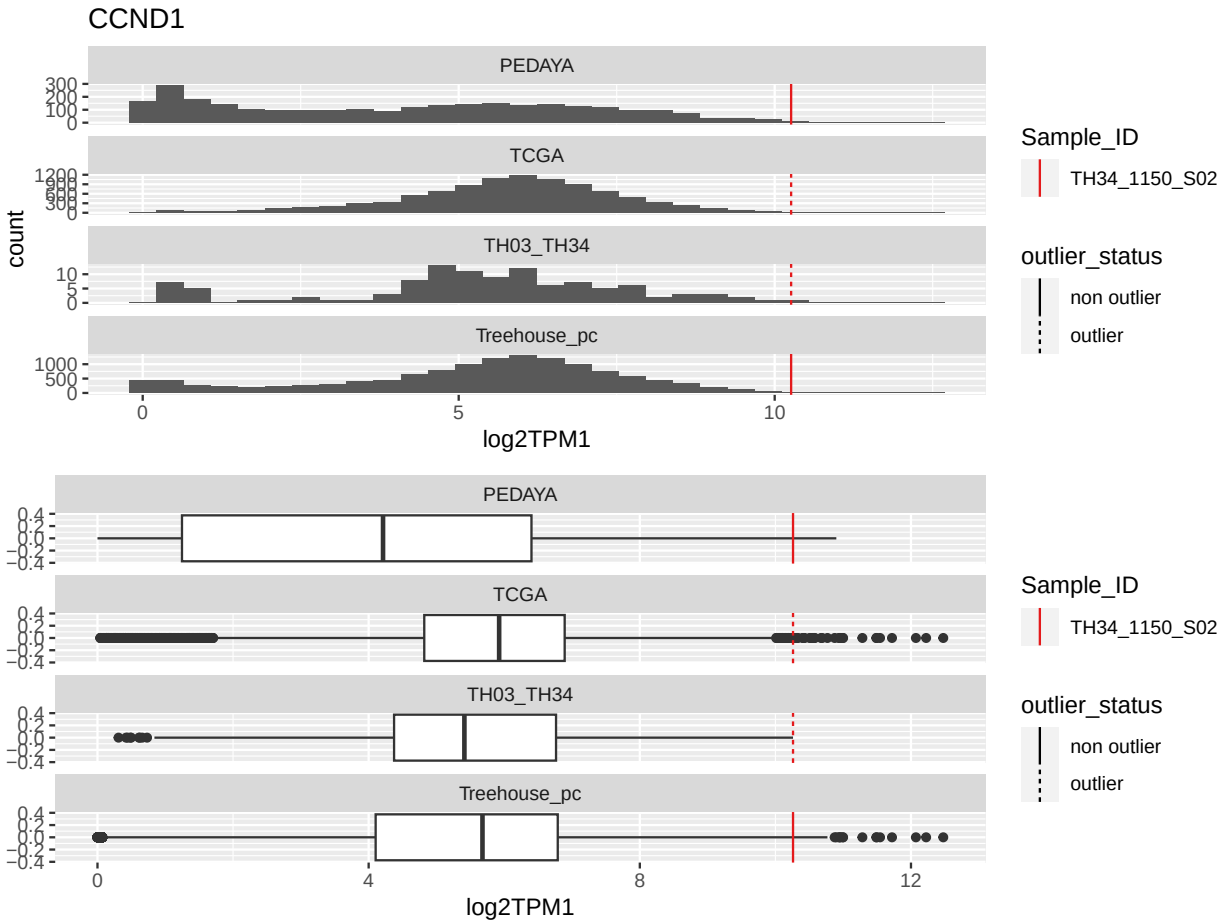
```
##
## [[5]]
```



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

cohort	q25	median	q75	IQR	up_outlier_threshold
PEDAYA	1.42	2.79	5.00	3.57	10.36
TCGA	1.22	1.98	2.82	1.60	5.21
TH03_TH34	1.31	2.31	3.69	2.37	7.25
Treehouse_pc	1.22	2.08	3.08	1.86	5.87

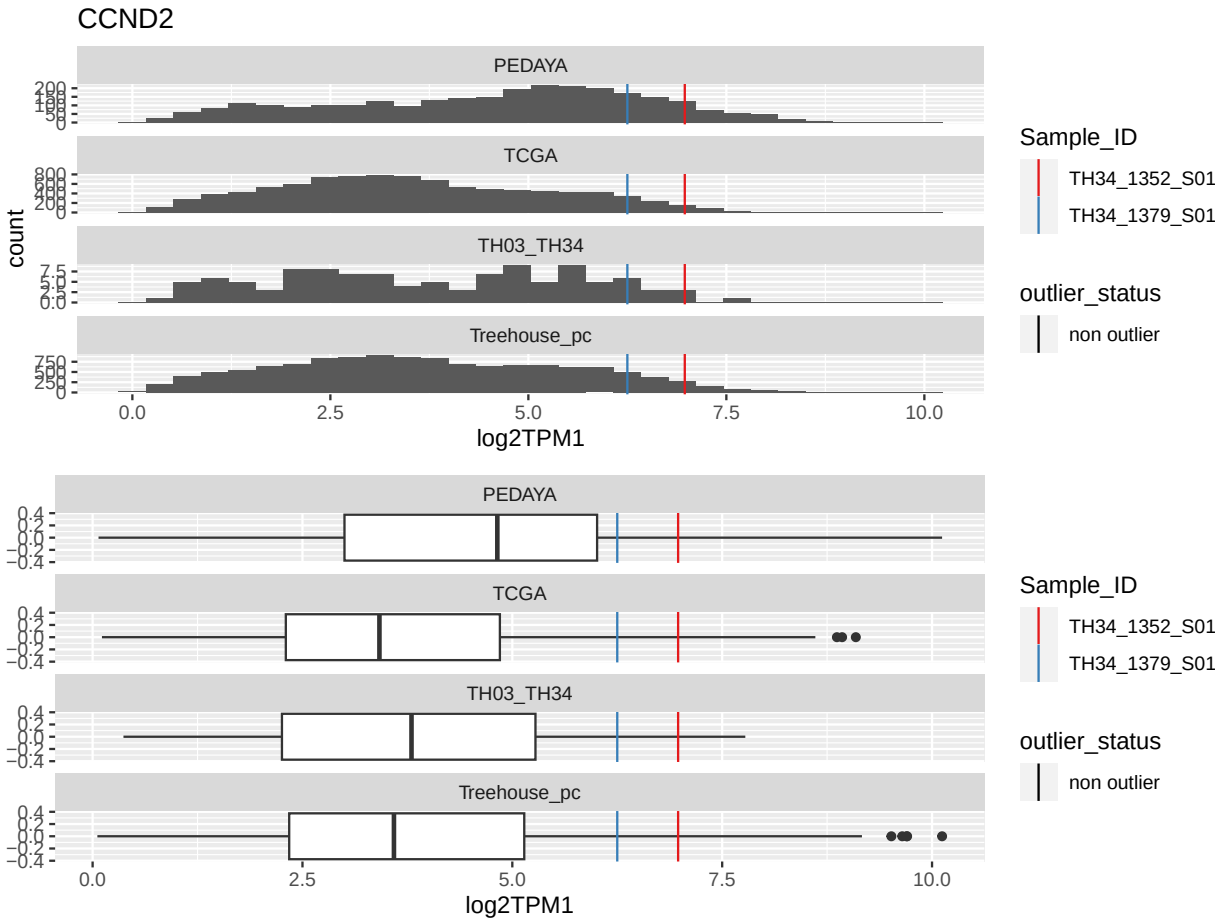
```
##
## [[6]]
```



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

cohort	q25	median	q75	IQR	up_outlier_threshold
PEDAYA	1.25	4.21	6.40	5.16	14.14
TCGA	4.82	5.92	6.89	2.07	10.00
TH03_TH34	4.37	5.41	6.76	2.39	10.35
Treehouse_pc	4.10	5.68	6.79	2.69	10.82

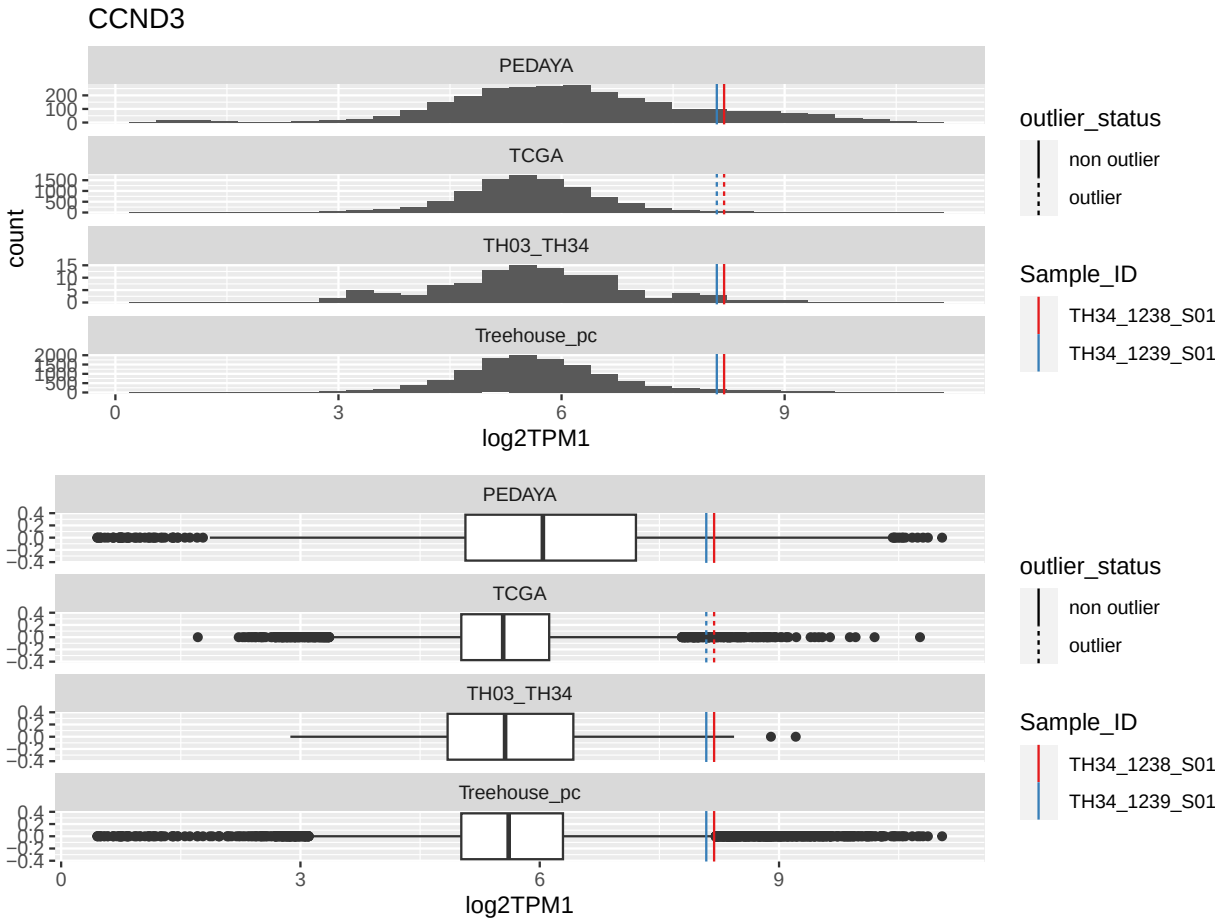
```
##
## [[7]]
```



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

cohort	q25	median	q75	IQR	up_outlier_threshold
PEDAYA	3.00	4.82	6.01	3.01	10.52
TCGA	2.30	3.42	4.85	2.55	8.68
TH03_TH34	2.26	3.80	5.28	3.02	9.81
Treehouse_pc	2.34	3.59	5.14	2.80	9.34

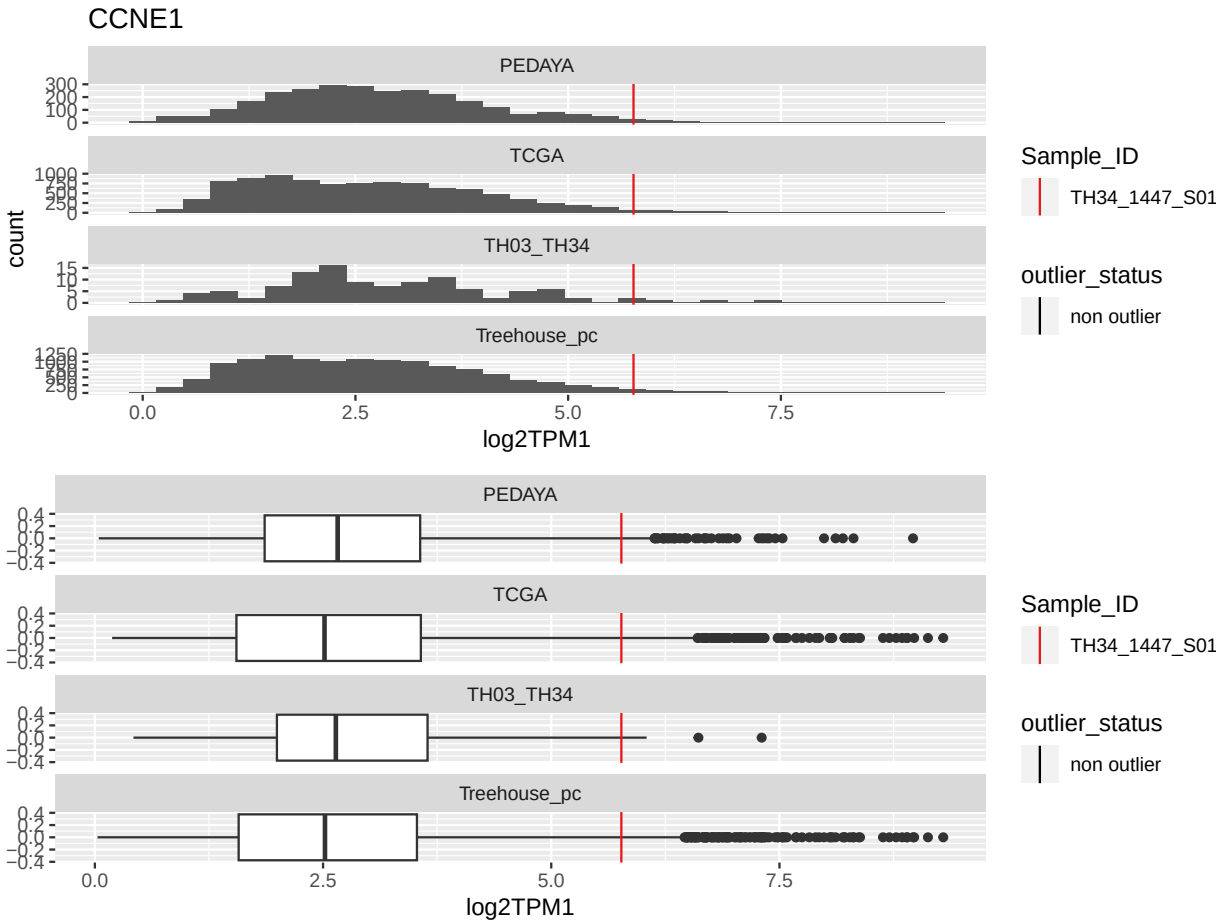
```
##
## [[8]]
```



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

cohort	q25	median	q75	IQR	up_outlier_threshold
PEDAYA	5.07	6.04	7.21	2.14	10.41
TCGA	5.02	5.54	6.12	1.10	7.77
TH03_TH34	4.84	5.57	6.42	1.58	8.79
Treehouse_pc	5.02	5.61	6.29	1.27	8.20

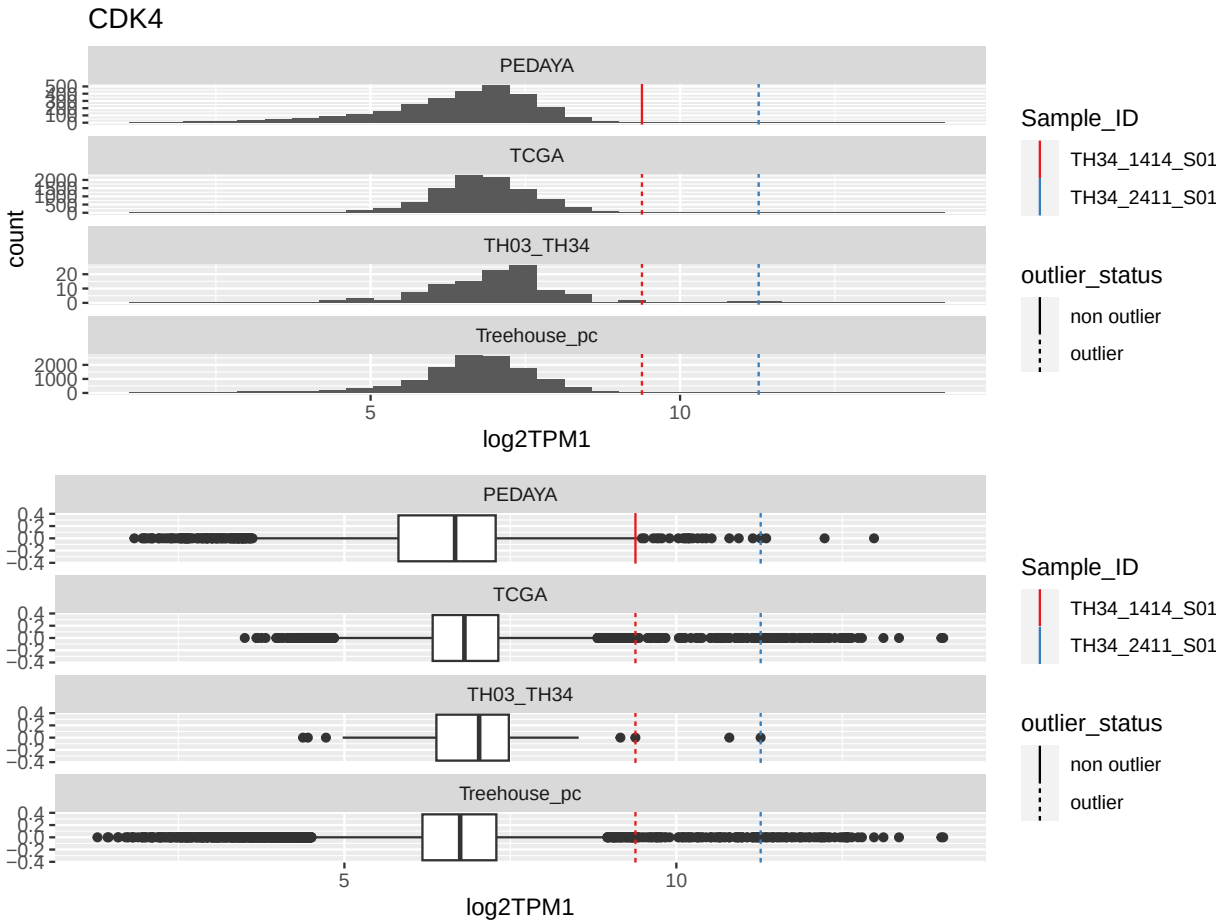
```
##
## [[9]]
```



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

cohort	q25	median	q75	IQR	up_outlier_threshold
PEDAYA	1.86	2.66	3.56	1.70	6.12
TCGA	1.55	2.52	3.57	2.02	6.60
TH03_TH34	1.99	2.64	3.64	1.65	6.12
Treehouse_pc	1.58	2.52	3.53	1.95	6.46

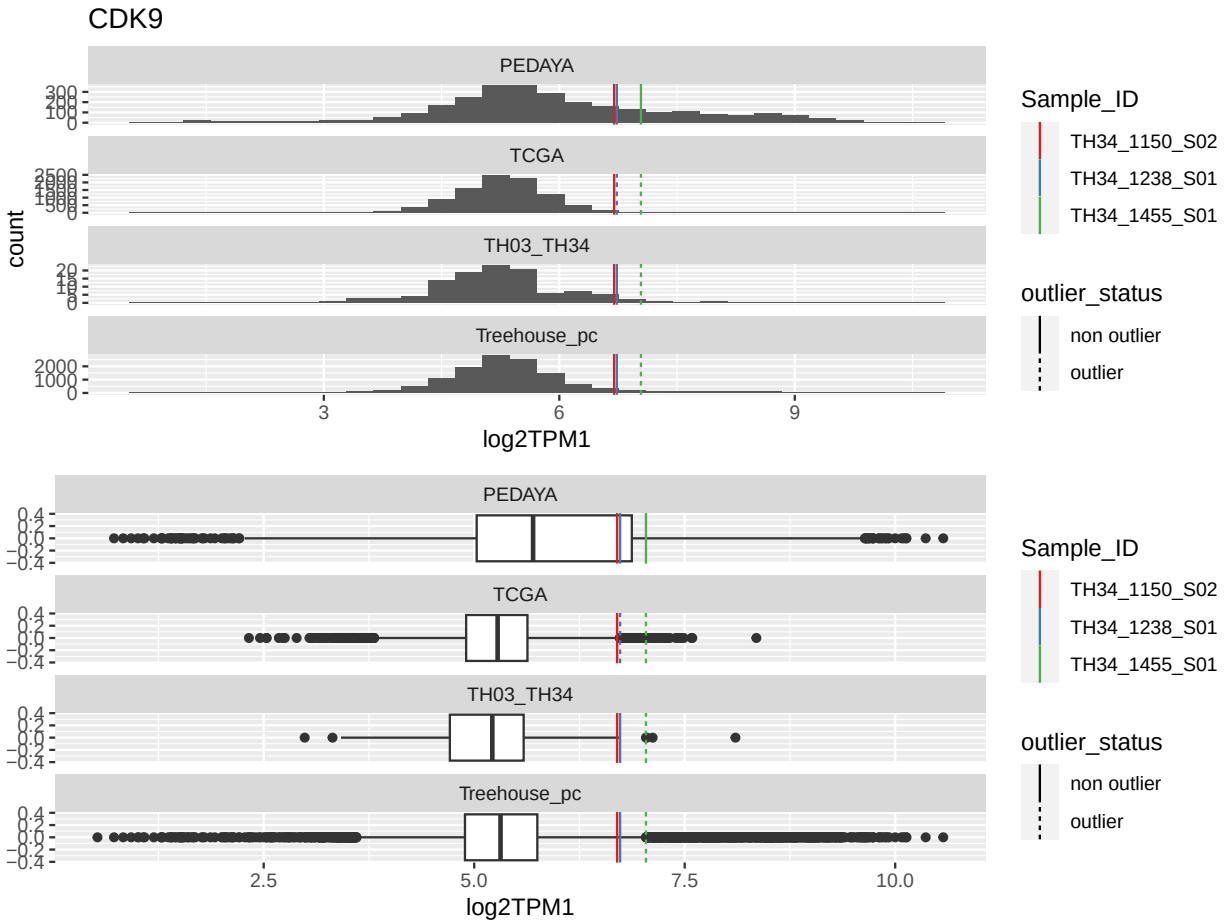
```
##
## [[10]]
```



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

cohort	q25	median	q75	IQR	up_outlier_threshold
PEDAYA	5.81	6.67	7.28	1.46	9.48
TCGA	6.33	6.81	7.32	0.99	8.80
TH03_TH34	6.39	7.03	7.48	1.09	9.11
Treehouse_pc	6.18	6.74	7.29	1.11	8.96

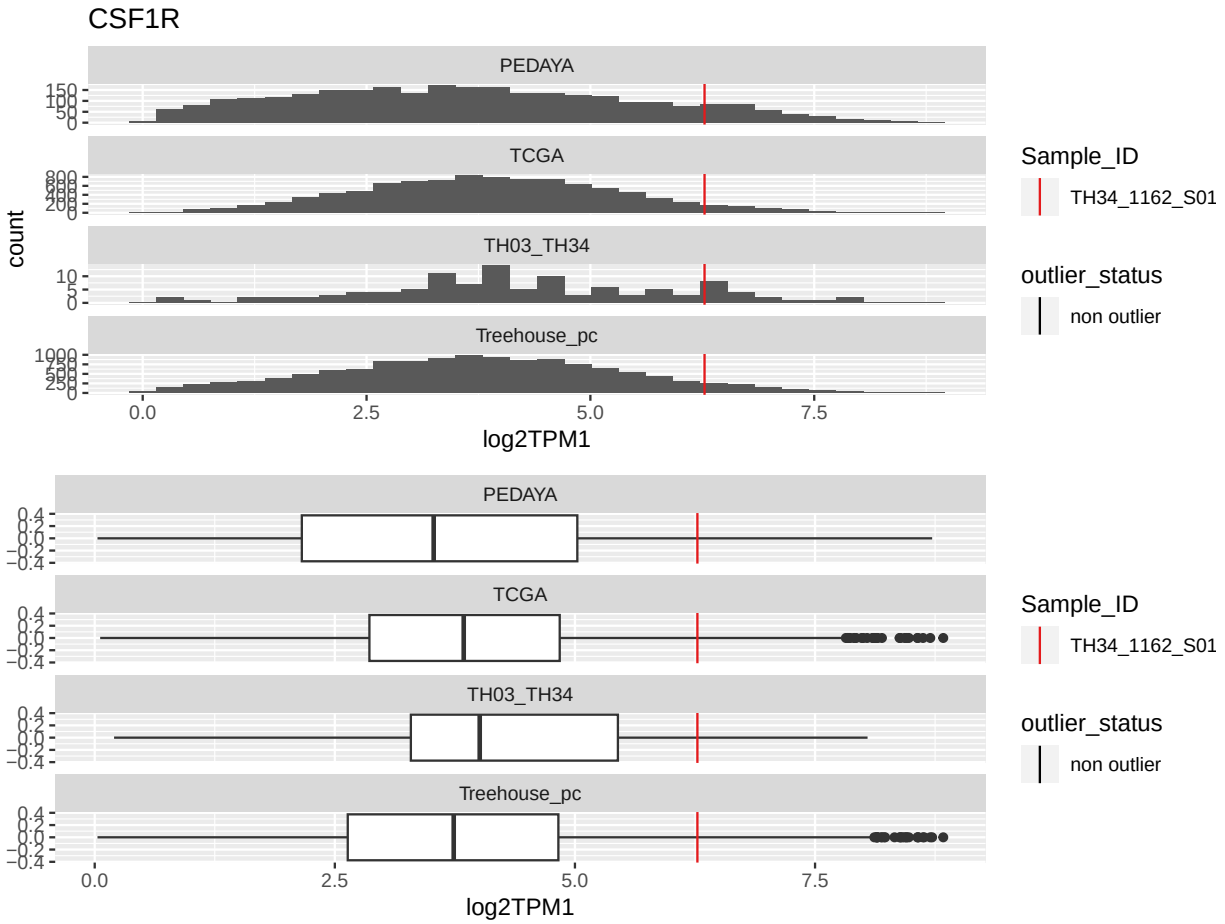
```
##
## [[11]]
```



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

cohort	q25	median	q75	IQR	up_outlier_threshold
PEDAYA	5.03	5.70	6.87	1.84	9.64
TCGA	4.90	5.28	5.63	0.73	6.72
TH03_TH34	4.71	5.22	5.59	0.88	6.90
Treehouse_pc	4.89	5.31	5.75	0.86	7.04

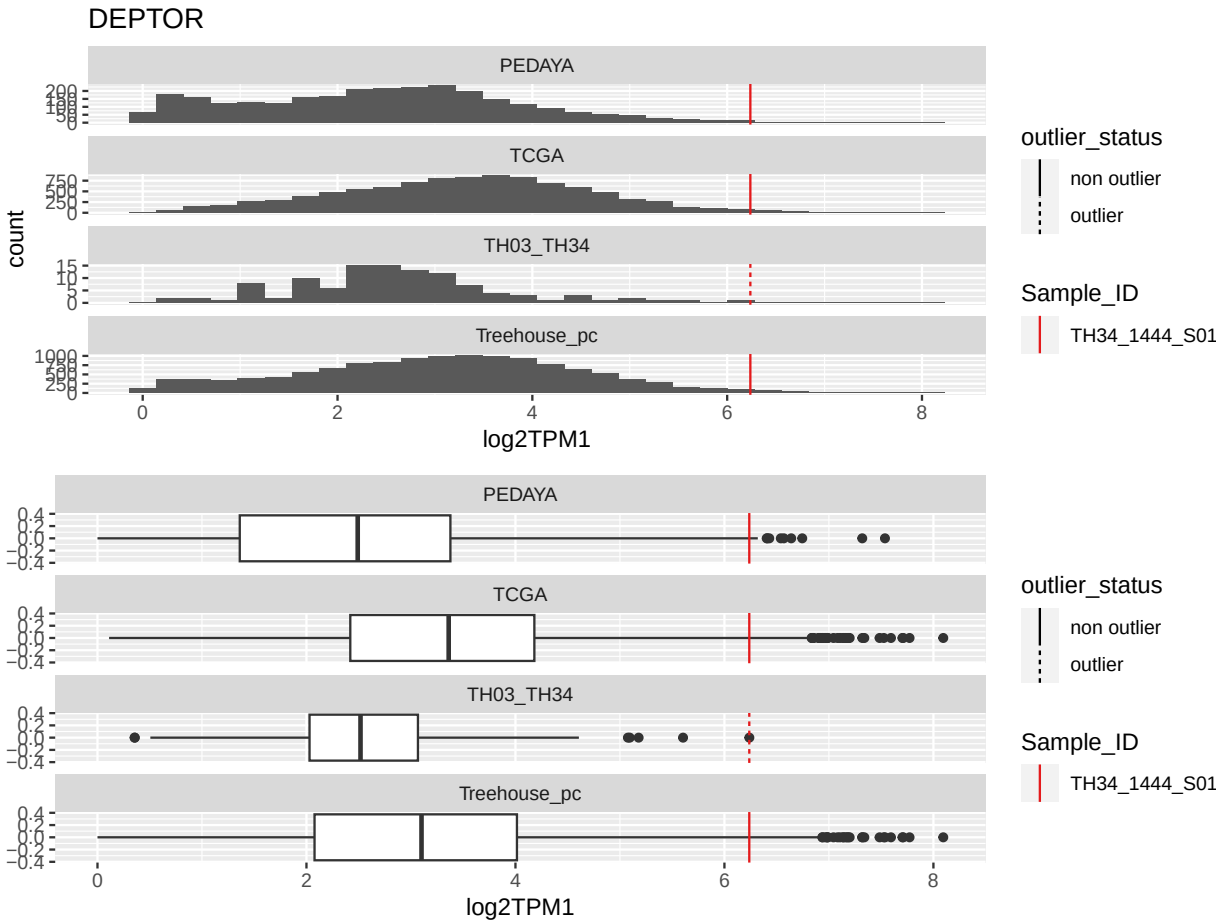
```
##
## [[12]]
```



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

cohort	q25	median	q75	IQR	up_outlier_threshold
PEDAYA	2.16	3.53	5.02	2.87	9.33
TCGA	2.86	3.84	4.84	1.98	7.81
TH03_TH34	3.29	4.01	5.45	2.16	8.68
Treehouse_pc	2.63	3.74	4.83	2.19	8.12

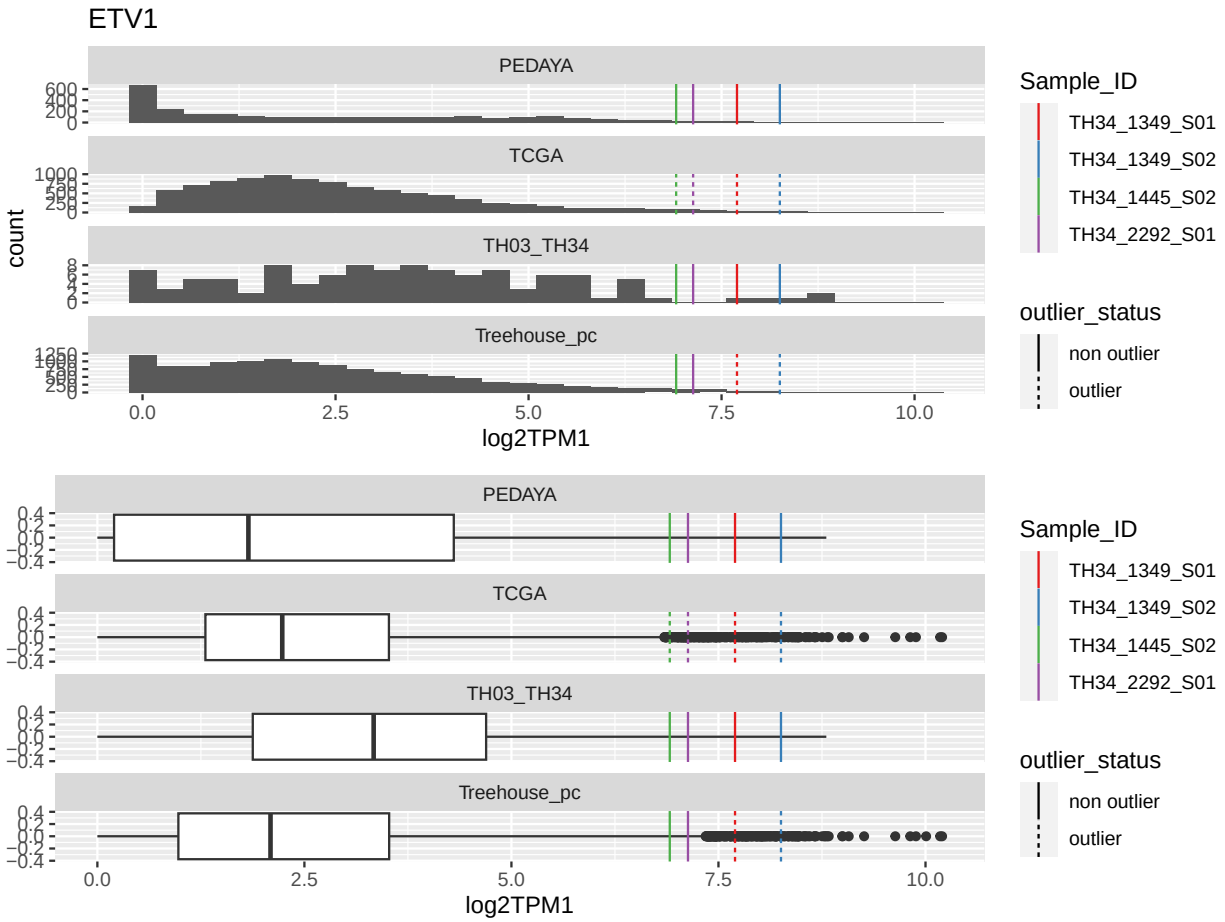
```
##
## [[13]]
```



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

cohort	q25	median	q75	IQR	up_outlier_threshold
PEDAYA	1.36	2.49	3.38	2.02	6.40
TCGA	2.42	3.36	4.18	1.76	6.82
TH03_TH34	2.03	2.52	3.07	1.04	4.62
Treehouse_pc	2.08	3.10	4.01	1.94	6.92

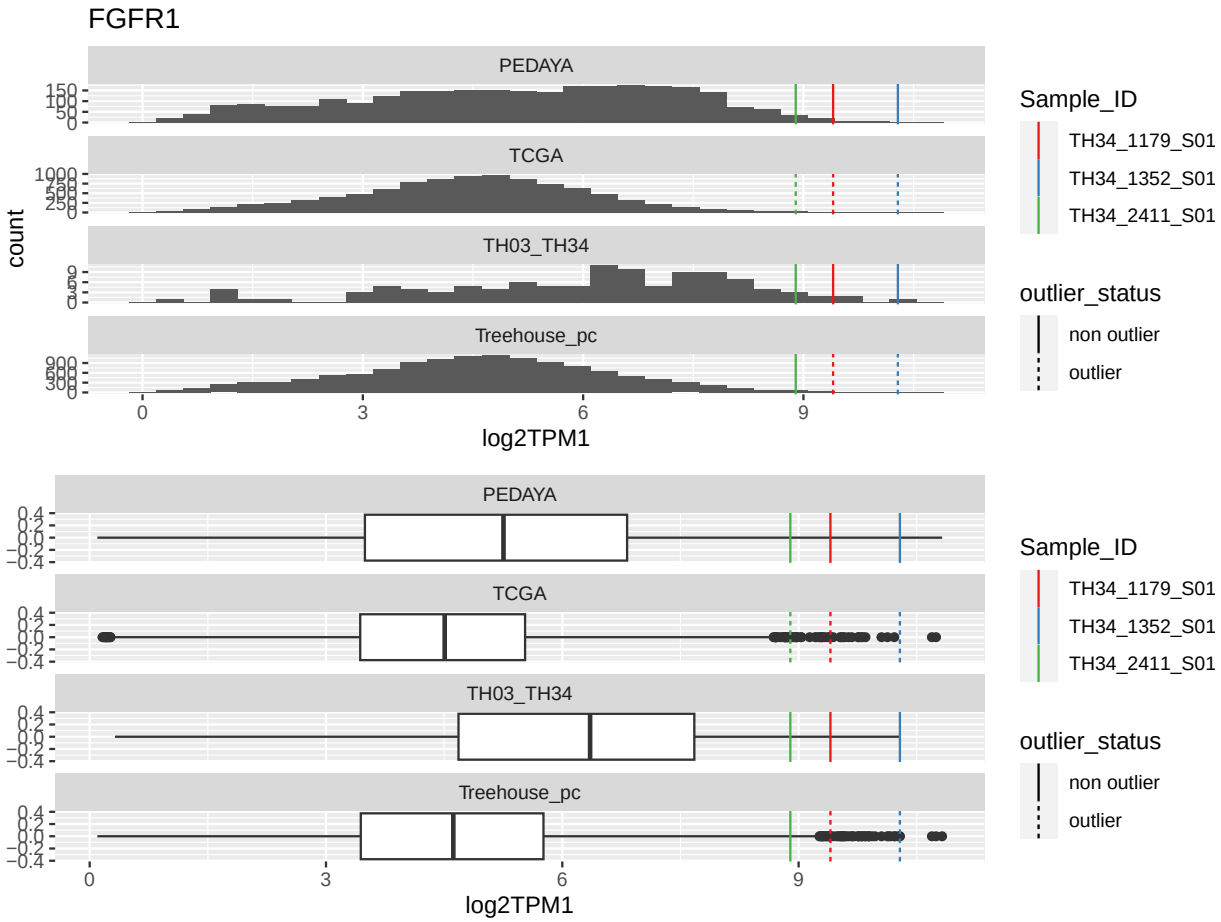
```
##
## [[14]]
```



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_2292_S01	ETV1	FALSE	TRUE	FALSE	FALSE	7.132

cohort	q25	median	q75	IQR	up_outlier_threshold
PEDAYA	0.20	1.82	4.30	4.10	10.46
TCGA	1.30	2.23	3.52	2.22	6.84
TH03_TH34	1.88	3.34	4.69	2.82	8.92
Treehouse_pc	0.98	2.09	3.52	2.55	7.34

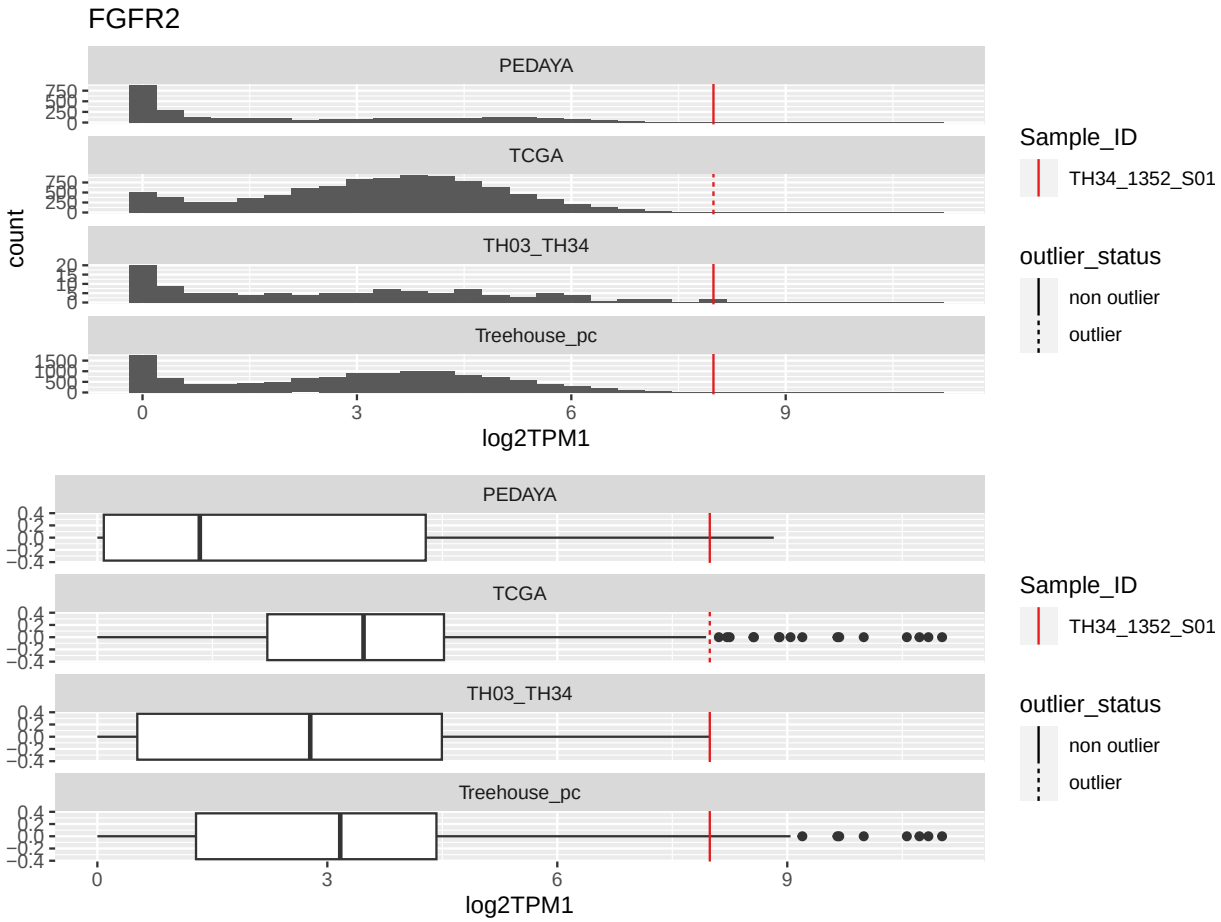
```
##
## [[15]]
```



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

cohort	q25	median	q75	IQR	up_outlier_threshold
PEDAYA	3.49	5.25	6.82	3.33	11.81
TCGA	3.43	4.51	5.53	2.09	8.67
TH03_TH34	4.68	6.35	7.68	2.99	12.17
Treehouse_pc	3.44	4.62	5.76	2.32	9.24

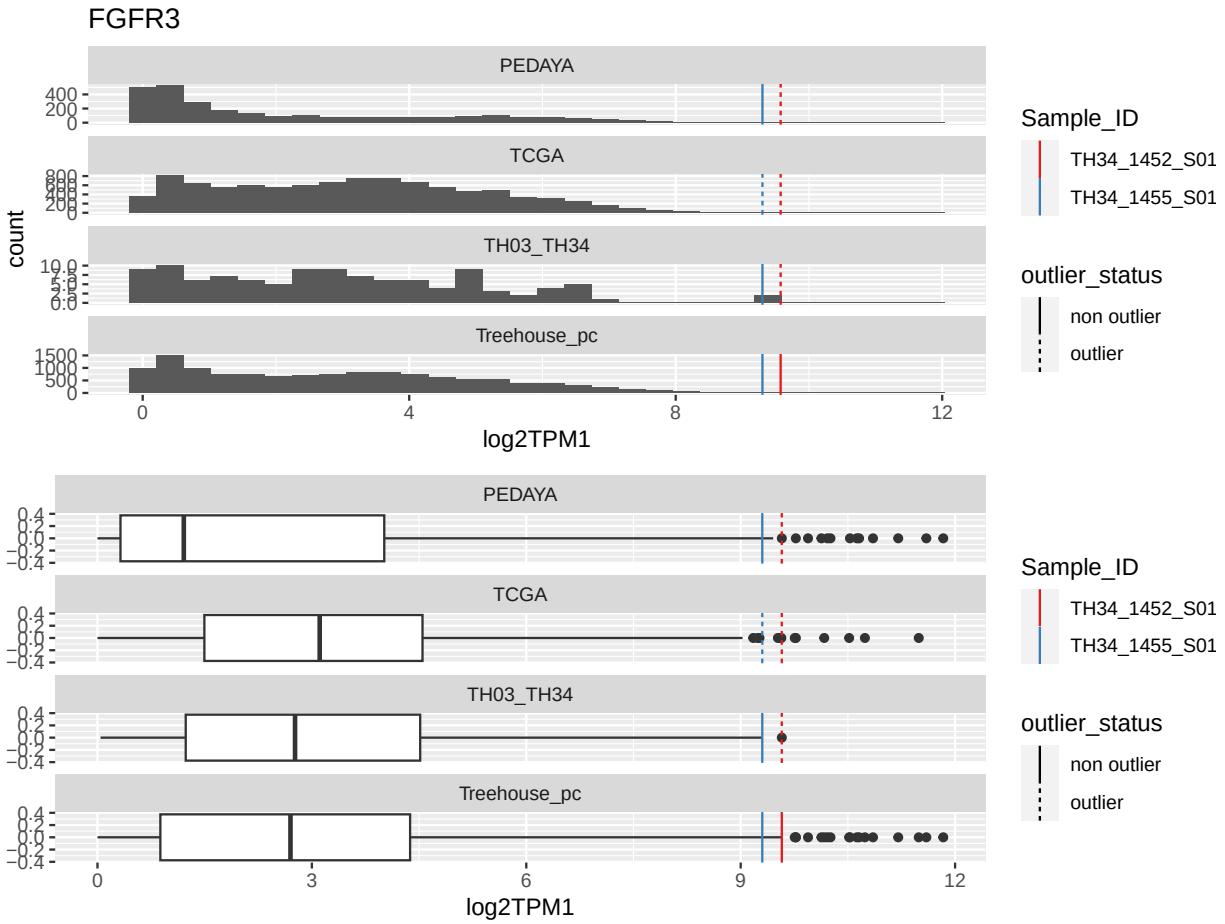
```
##
## [[16]]
```



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

cohort	q25	median	q75	IQR	up_outlier_threshold
PEDAYA	0.08	1.34	4.28	4.20	10.58
TCGA	2.22	3.47	4.52	2.30	7.98
TH03_TH34	0.52	2.78	4.49	3.97	10.45
Treehouse_pc	1.29	3.17	4.42	3.14	9.13

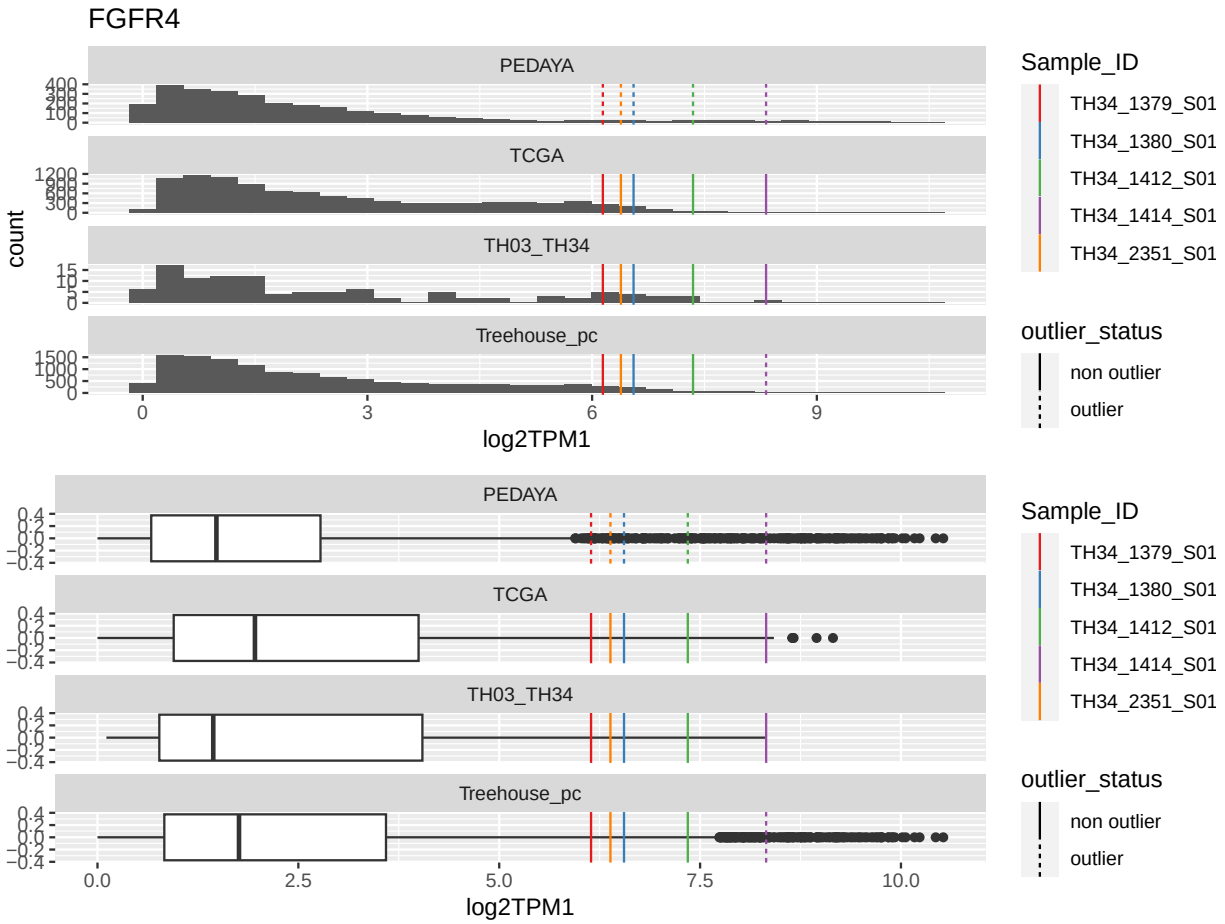
```
##
## [[17]]
```



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

cohort	q25	median	q75	IQR	up_outlier_threshold
PEDAYA	0.32	1.21	4.01	3.69	9.55
TCGA	1.50	3.11	4.55	3.05	9.13
TH03_TH34	1.23	2.76	4.52	3.28	9.44
Treehouse_pc	0.88	2.70	4.37	3.49	9.62

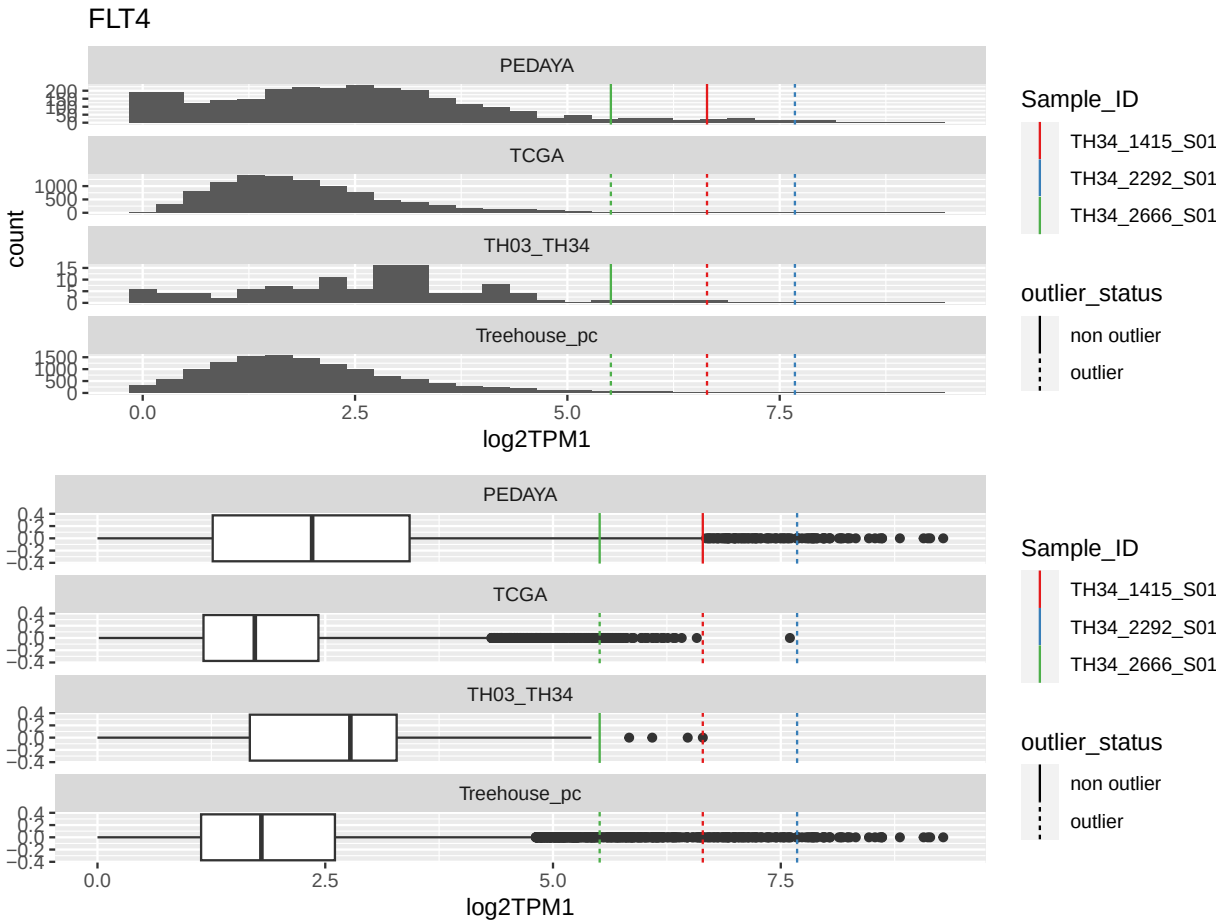
```
##
## [[18]]
```



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

cohort	q25	median	q75	IQR	up_outlier_threshold
PEDAYA	0.67	1.48	2.78	2.11	5.94
TCGA	0.95	1.96	4.00	3.05	8.57
TH03_TH34	0.77	1.44	4.04	3.27	8.96
Treehouse_pc	0.83	1.76	3.59	2.76	7.73

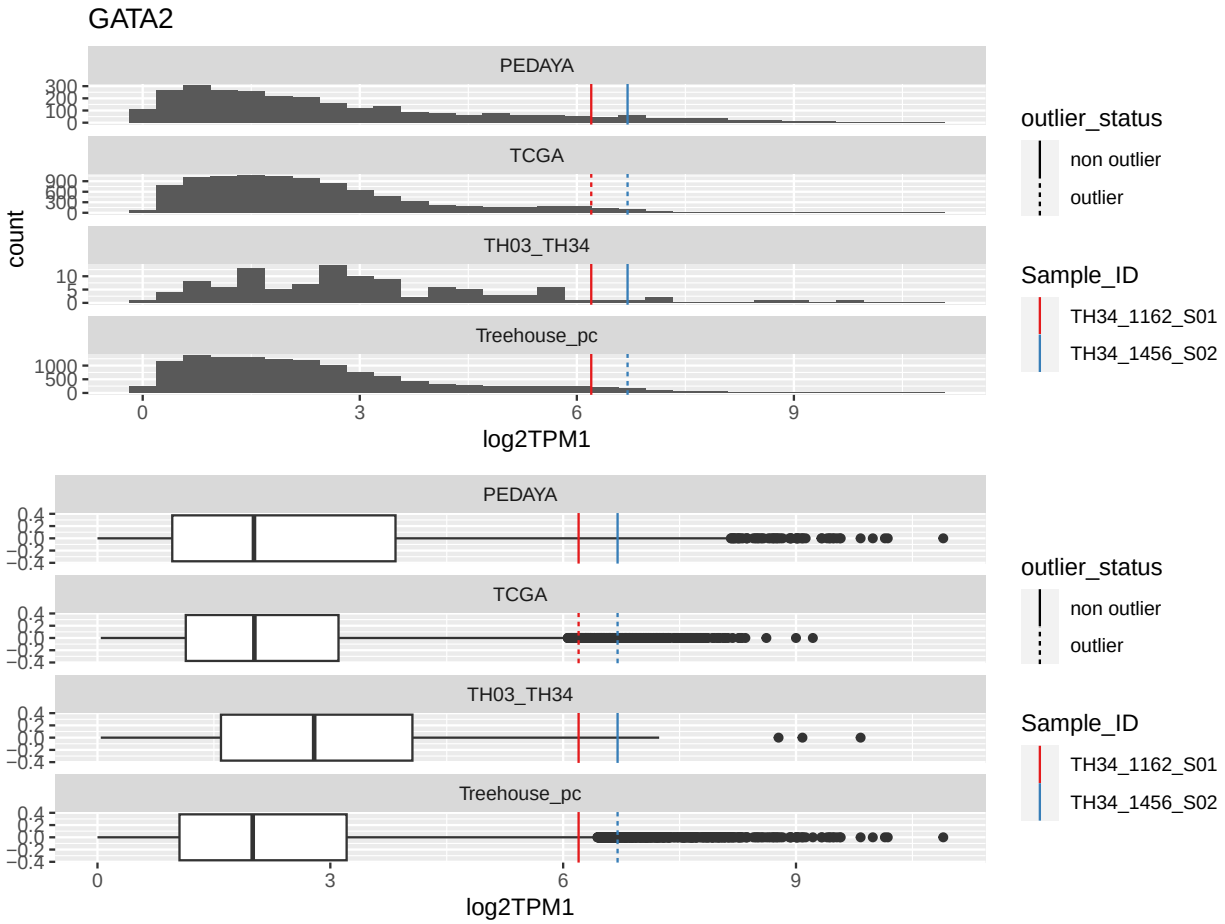
```
##
## [[19]]
```



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

cohort	q25	median	q75	IQR	up_outlier_threshold
PEDAYA	1.26	2.36	3.43	2.16	6.67
TCGA	1.16	1.73	2.42	1.26	4.32
TH03_TH34	1.67	2.78	3.28	1.61	5.70
Treehouse_pc	1.14	1.80	2.61	1.47	4.81

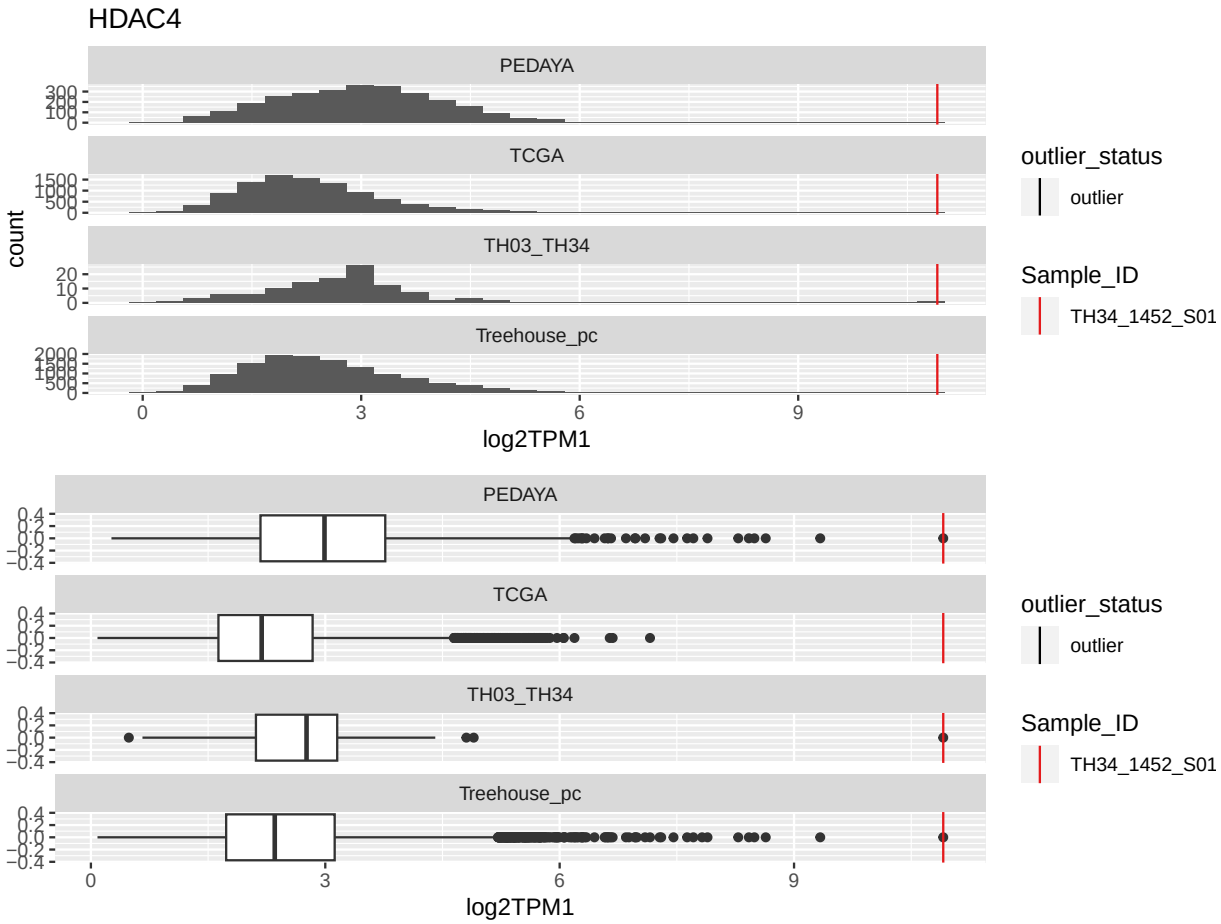
```
##
## [[20]]
```



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

cohort	q25	median	q75	IQR	up_outlier_threshold
PEDAYA	0.96	2.02	3.84	2.88	8.15
TCGA	1.14	2.02	3.11	1.97	6.06
TH03_TH34	1.59	2.79	4.06	2.47	7.76
Treehouse_pc	1.06	2.00	3.21	2.15	6.44

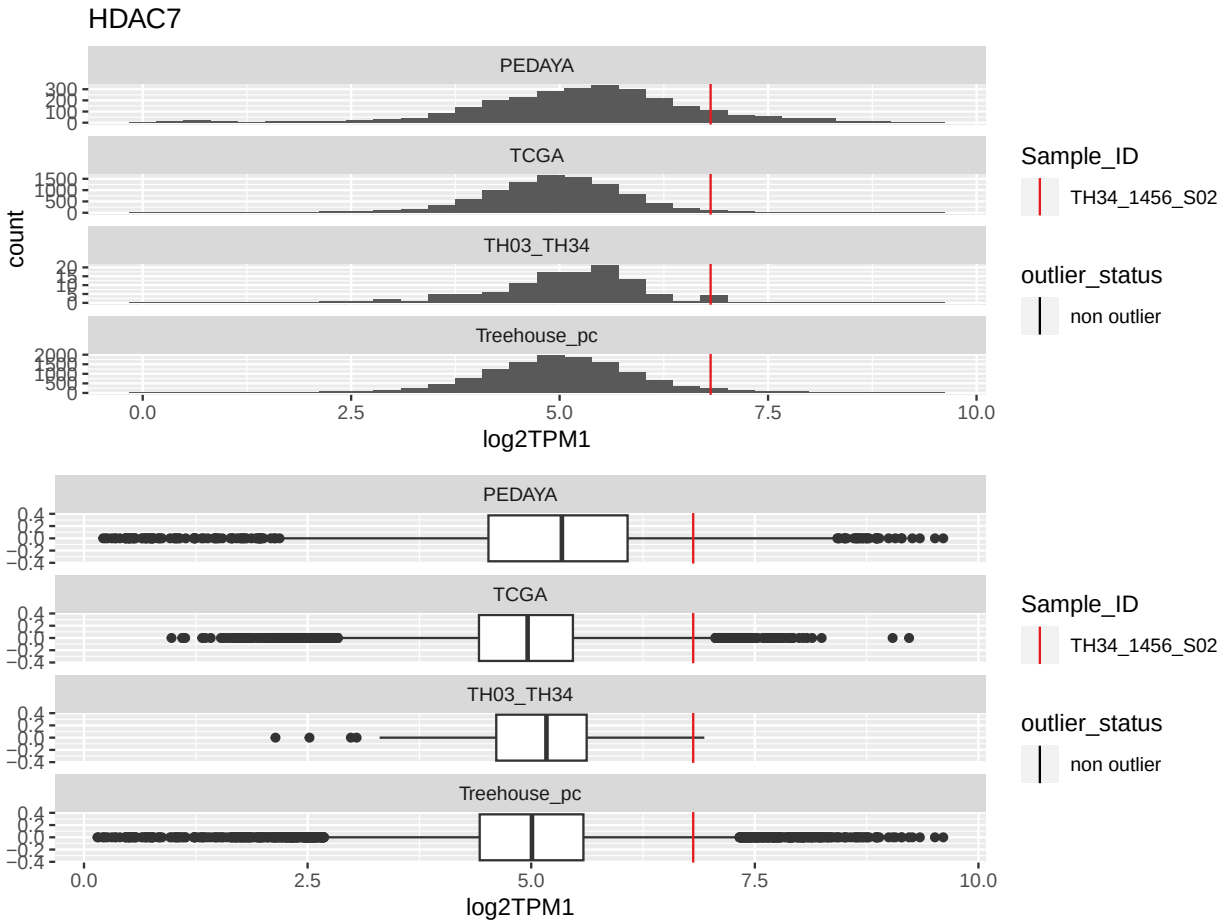
```
##
## [[21]]
```



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

cohort	q25	median	q75	IQR	up_outlier_threshold
PEDAYA	2.17	2.99	3.77	1.60	6.16
TCGA	1.63	2.19	2.84	1.21	4.65
TH03_TH34	2.11	2.76	3.15	1.04	4.71
Treehouse_pc	1.73	2.35	3.12	1.39	5.20

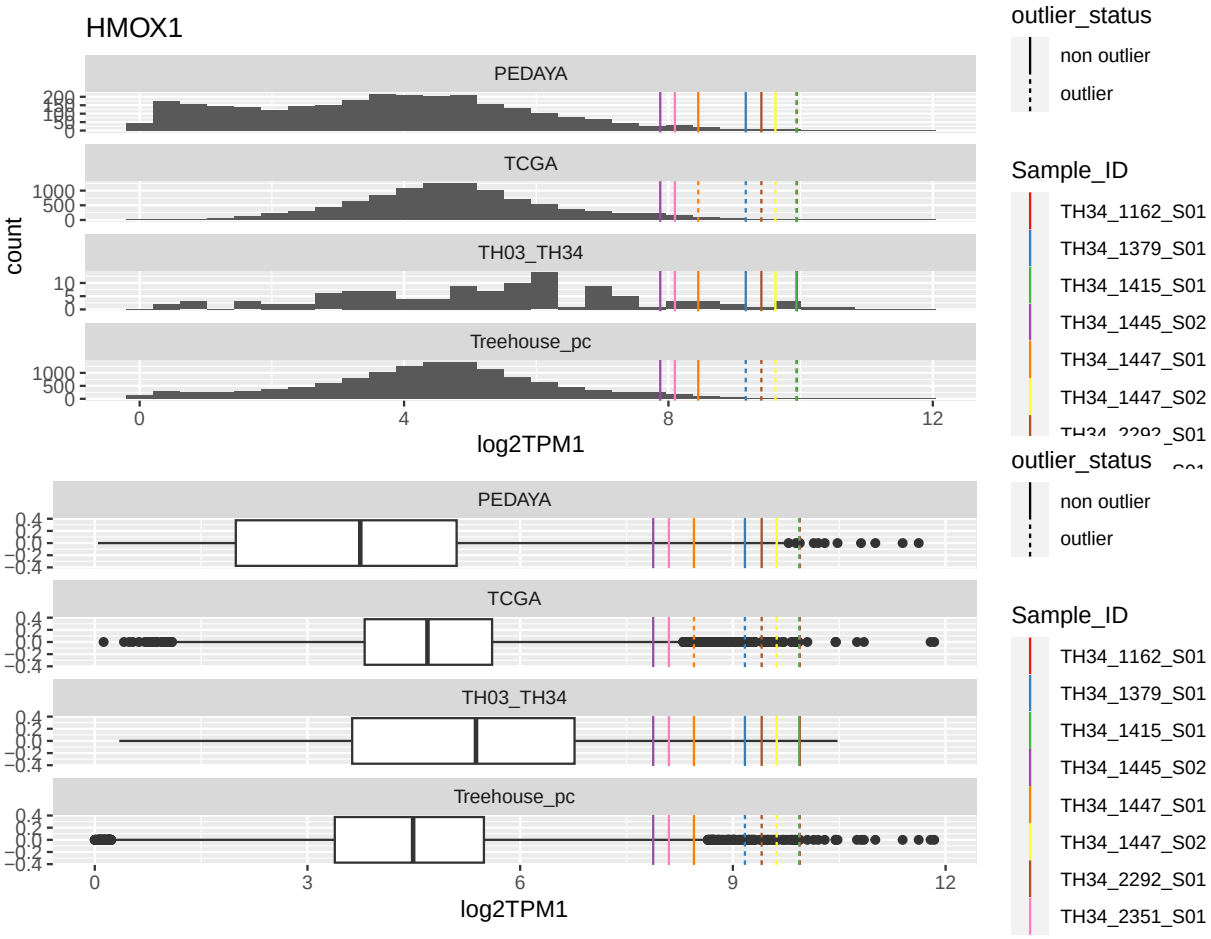
```
##
## [[22]]
```



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

cohort	q25	median	q75	IQR	up_outlier_threshold
PEDAYA	4.52	5.34	6.08	1.56	8.41
TCGA	4.42	4.96	5.47	1.05	7.04
TH03_TH34	4.61	5.17	5.62	1.01	7.14
Treehouse_pc	4.42	5.01	5.58	1.16	7.32

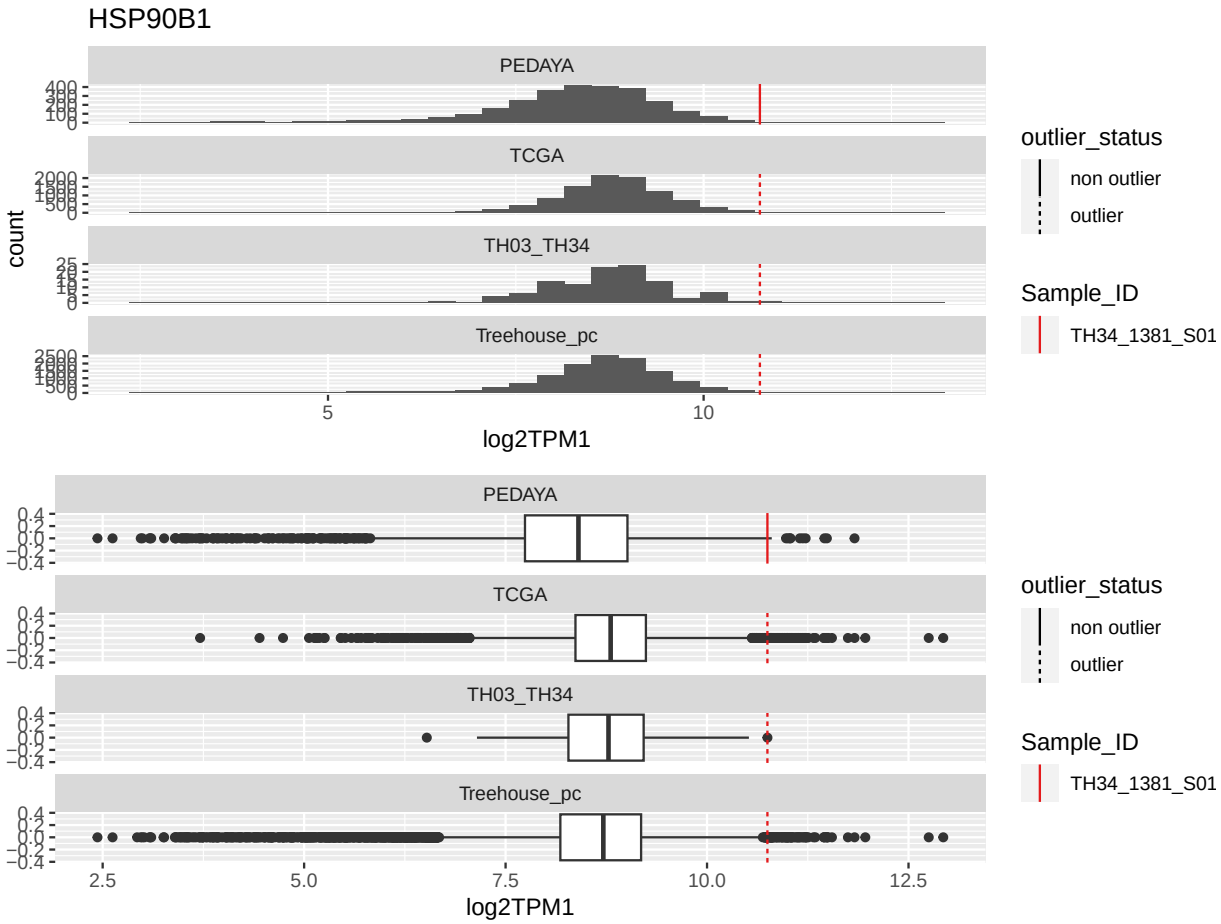
```
##
## [[23]]
```



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_1445_S02	HMOX1	FALSE	FALSE	FALSE	FALSE	7.876
TH34_2351_S01	HMOX1	FALSE	FALSE	FALSE	FALSE	8.098

cohort	q25	median	q75	IQR	up_outlier_threshold
PEDAYA	1.99	3.74	5.10	3.11	9.78
TCGA	3.81	4.69	5.60	1.80	8.30
TH03_TH34	3.63	5.38	6.77	3.14	11.47
Treehouse_pc	3.39	4.49	5.49	2.10	8.64

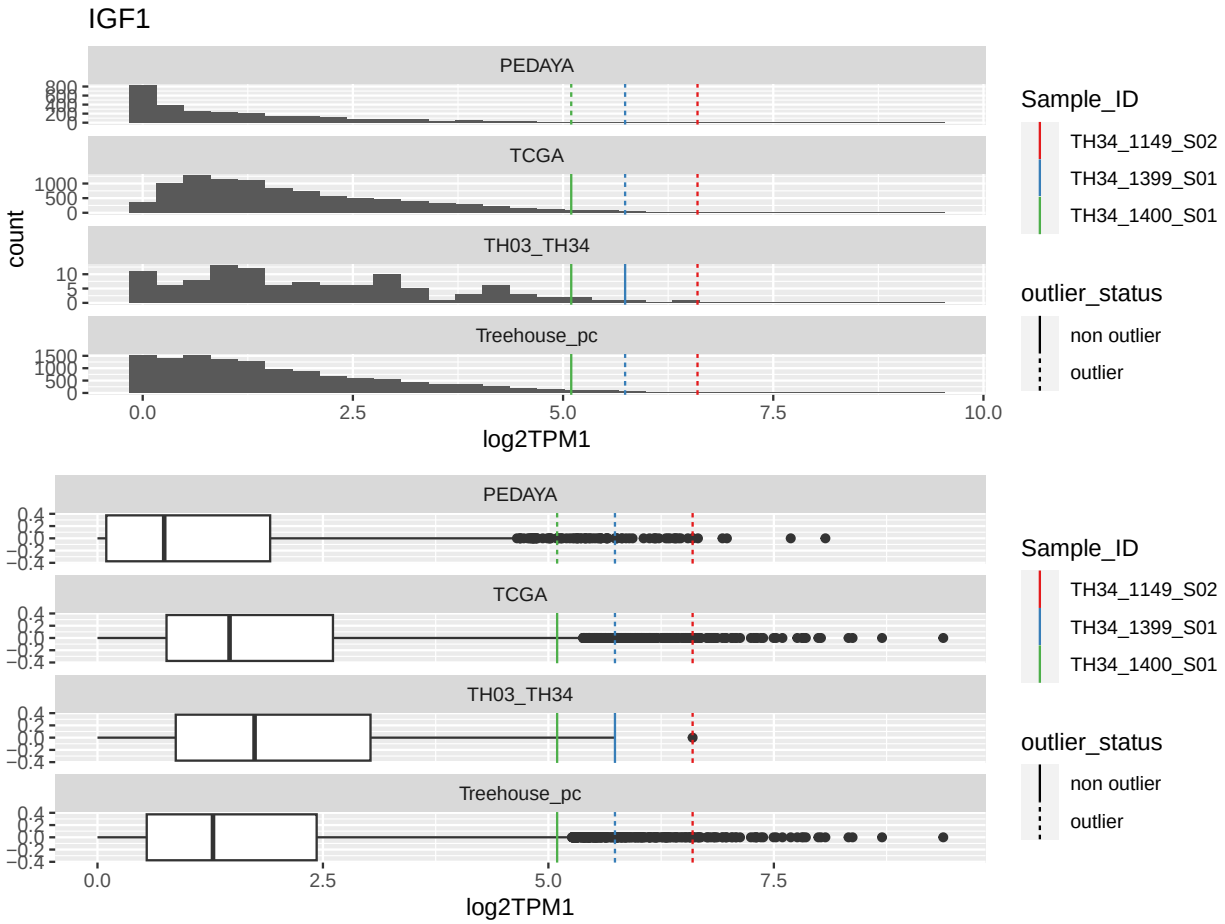
```
##
## [[24]]
```



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

cohort	q25	median	q75	IQR	up_outlier_threshold
PEDAYA	7.74	8.40	9.01	1.27	10.92
TCGA	8.37	8.80	9.24	0.87	10.55
TH03_TH34	8.28	8.78	9.21	0.93	10.61
Treehouse_pc	8.18	8.71	9.18	1.00	10.68

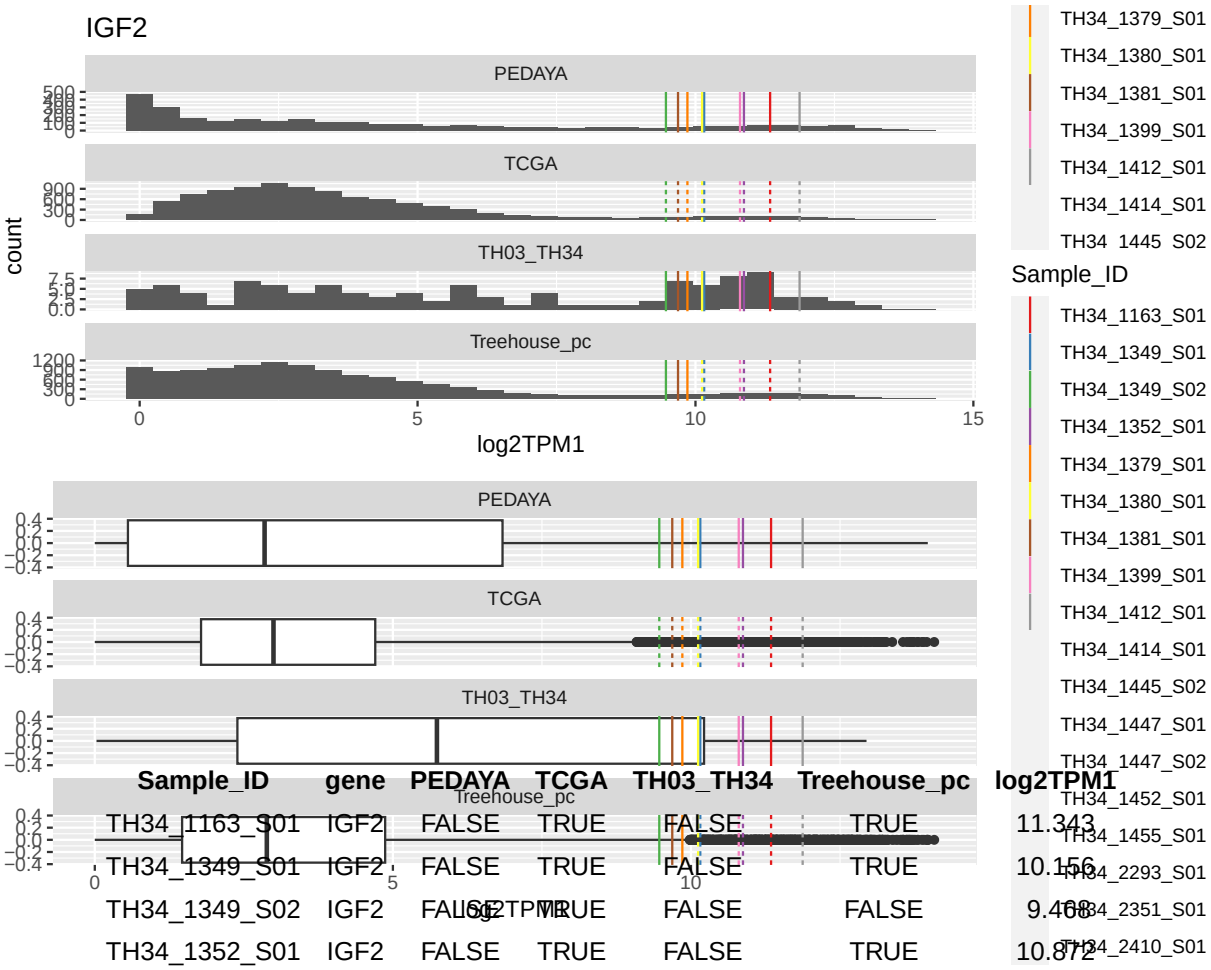
```
##
## [[25]]
```



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

cohort	q25	median	q75	IQR	up_outlier_threshold
PEDAYA	0.10	0.74	1.91	1.82	4.64
TCGA	0.77	1.46	2.61	1.85	5.38
TH03_TH34	0.87	1.74	3.03	2.16	6.27
Treehouse_pc	0.55	1.28	2.43	1.88	5.26

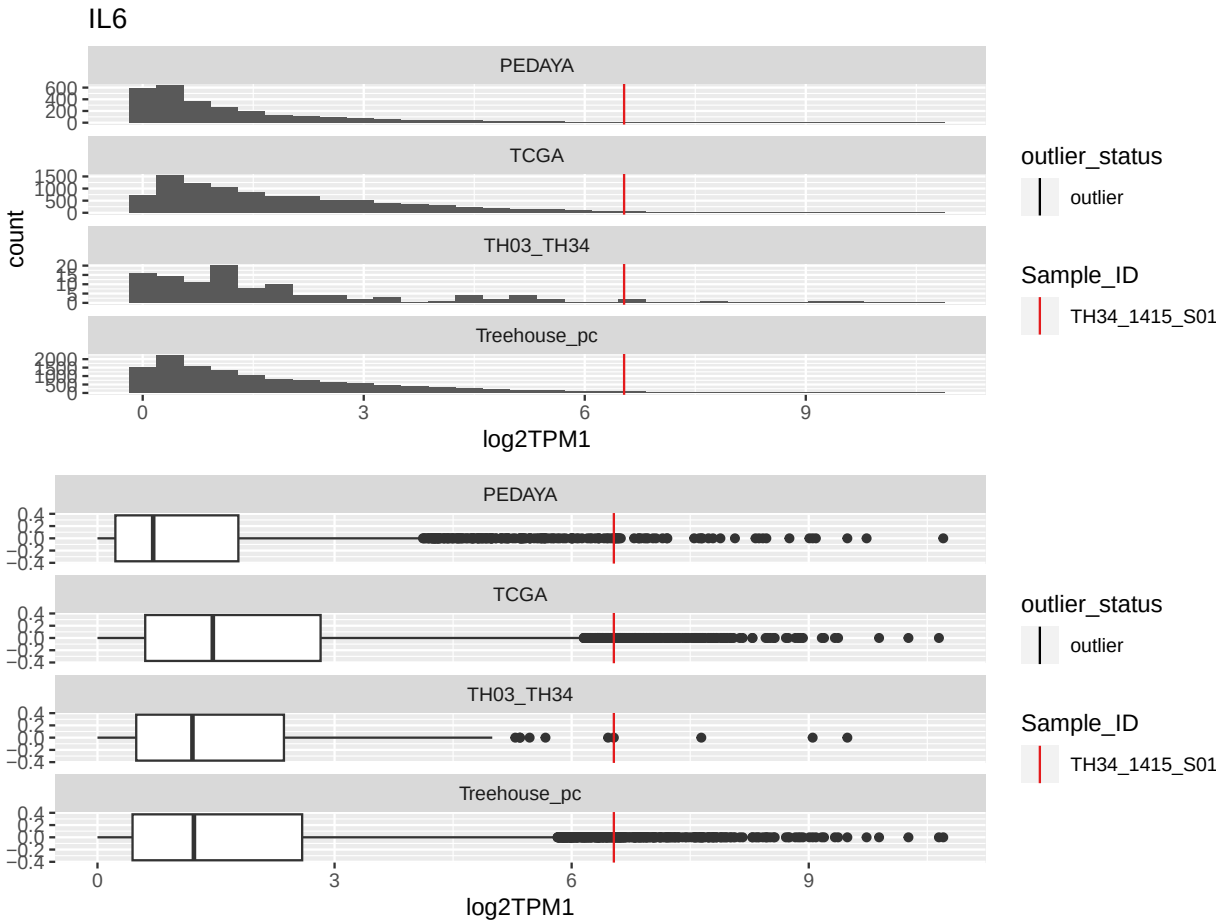
```
##
## [[26]]
```



Sample_ID	gene	PEDAYA	TCGA	TH03_TH34	Treehouse_pc	log2TPM1
TH34_1163_S01	IGF2	FALSE	TRUE	FALSE	TRUE	11.343
TH34_1349_S01	IGF2	FALSE	TRUE	FALSE	TRUE	10.156
TH34_1349_S02	IGF2	FALSE	TRUE	FALSE	FALSE	9.468
TH34_1352_S01	IGF2	FALSE	TRUE	FALSE	TRUE	10.872
TH34_1379_S01	IGF2	FALSE	TRUE	FALSE	FALSE	9.855
TH34_1380_S01	IGF2	FALSE	TRUE	FALSE	TRUE	10.119
TH34_1381_S01	IGF2	FALSE	TRUE	FALSE	FALSE	9.686
TH34_1399_S01	IGF2	FALSE	TRUE	FALSE	TRUE	10.801
TH34_1412_S01	IGF2	FALSE	TRUE	FALSE	TRUE	11.873
TH34_1414_S01	IGF2	FALSE	TRUE	FALSE	TRUE	10.477
TH34_1445_S02	IGF2	FALSE	TRUE	FALSE	FALSE	9.836
TH34_1447_S01	IGF2	FALSE	TRUE	FALSE	FALSE	9.719
TH34_1447_S02	IGF2	FALSE	TRUE	FALSE	TRUE	10.240
TH34_1452_S01	IGF2	FALSE	TRUE	FALSE	TRUE	10.642
TH34_1455_S01	IGF2	FALSE	TRUE	FALSE	TRUE	12.690
TH34_2293_S01	IGF2	FALSE	TRUE	FALSE	TRUE	10.012
TH34_2351_S01	IGF2	FALSE	TRUE	FALSE	TRUE	11.231
TH34_2410_S01	IGF2	FALSE	TRUE	FALSE	TRUE	9.884

cohort	q25	median	q75	IQR	up_outlier_threshold
PEDAYA	0.56	2.85	6.84	6.28	16.26
TCGA	1.78	3.00	4.70	2.92	9.09
TH03_TH34	2.39	5.74	10.22	7.83	21.96
Treehouse_pc	1.46	2.89	4.87	3.41	9.98

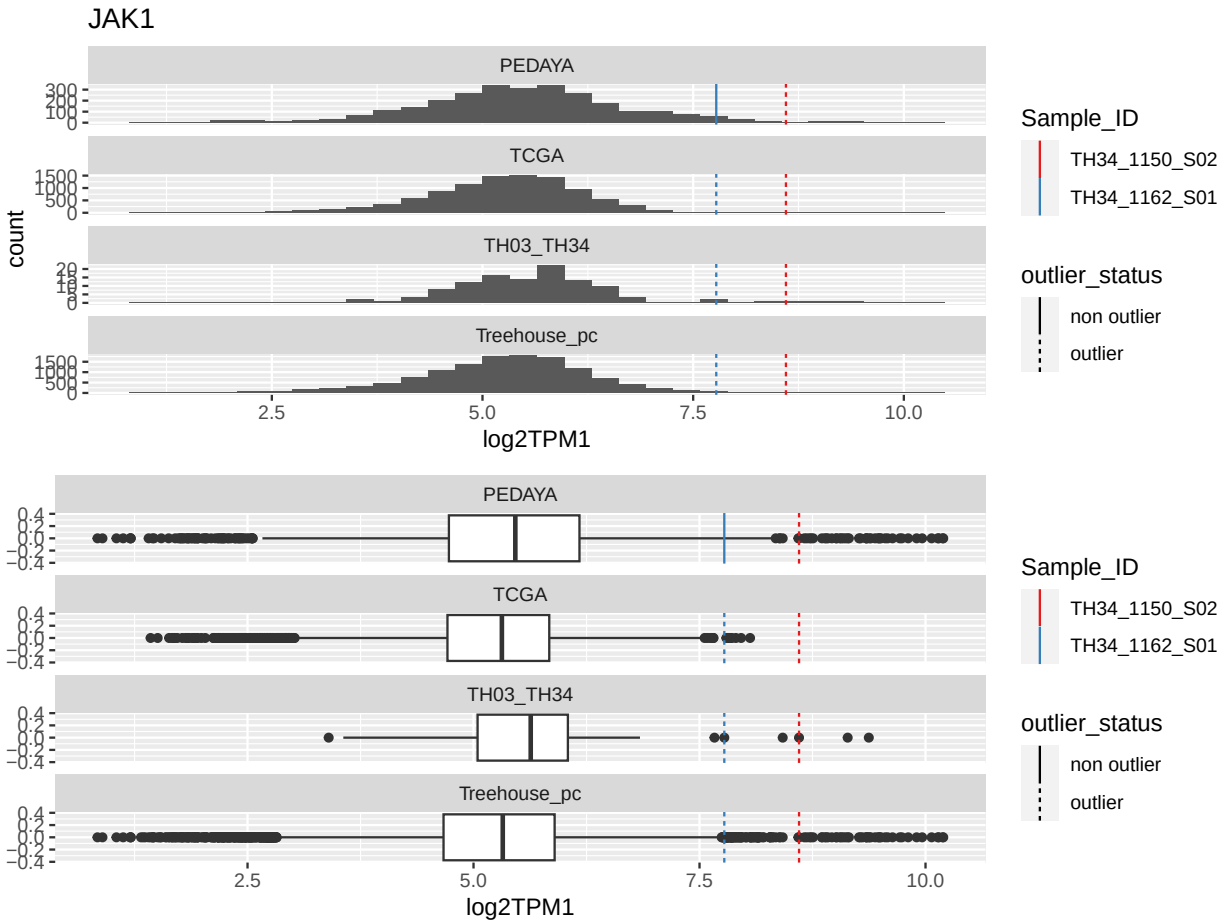
```
##
## [[27]]
```



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

cohort	q25	median	q75	IQR	up_outlier_threshold
PEDAYA	0.23	0.70	1.78	1.56	4.12
TCGA	0.60	1.46	2.82	2.22	6.15
TH03_TH34	0.49	1.20	2.36	1.87	5.16
Treehouse_pc	0.44	1.22	2.59	2.15	5.81

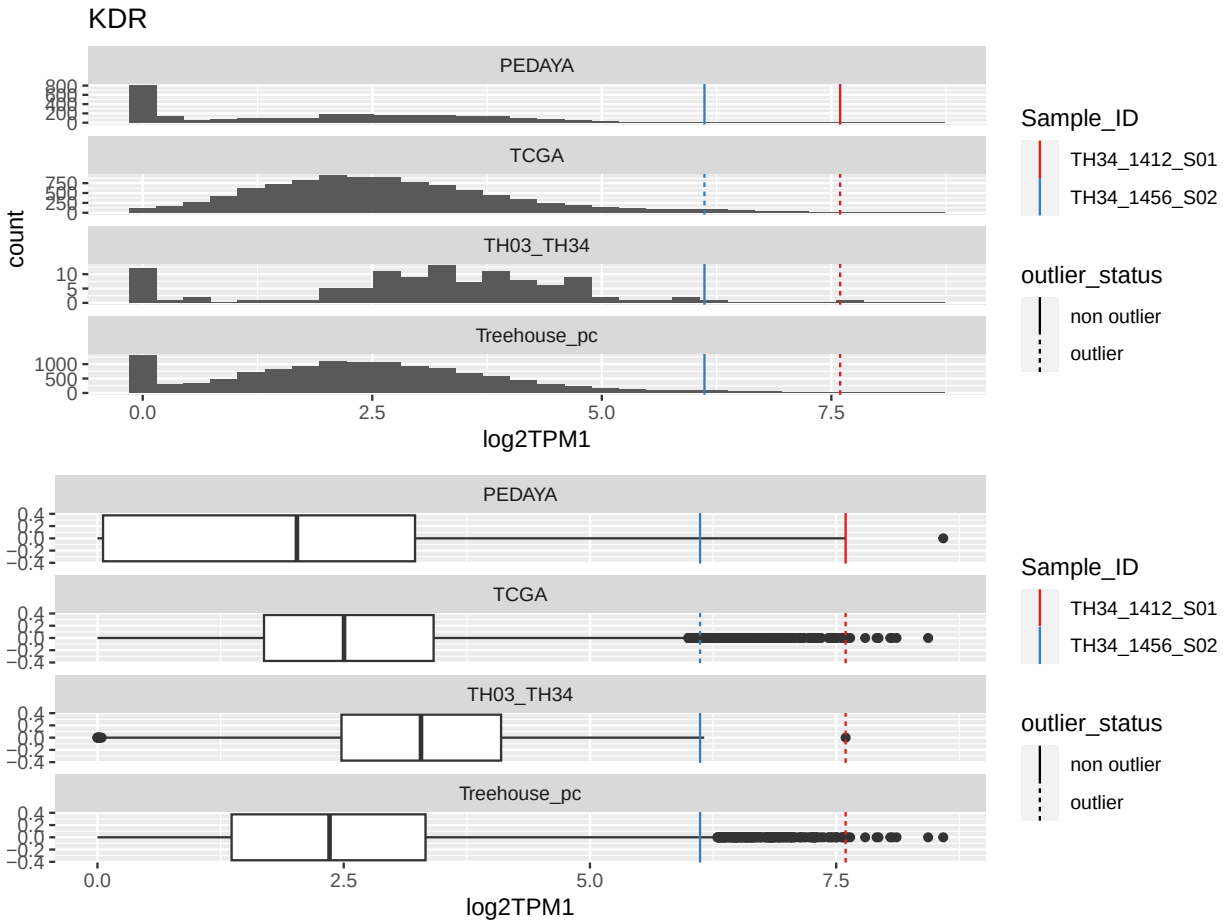
```
##
## [[28]]
```



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

cohort	q25	median	q75	IQR	up_outlier_threshold
PEDAYA	4.73	5.46	6.17	1.44	8.34
TCGA	4.71	5.31	5.84	1.13	7.53
TH03_TH34	5.04	5.63	6.04	1.00	7.54
Treehouse_pc	4.67	5.32	5.90	1.23	7.74

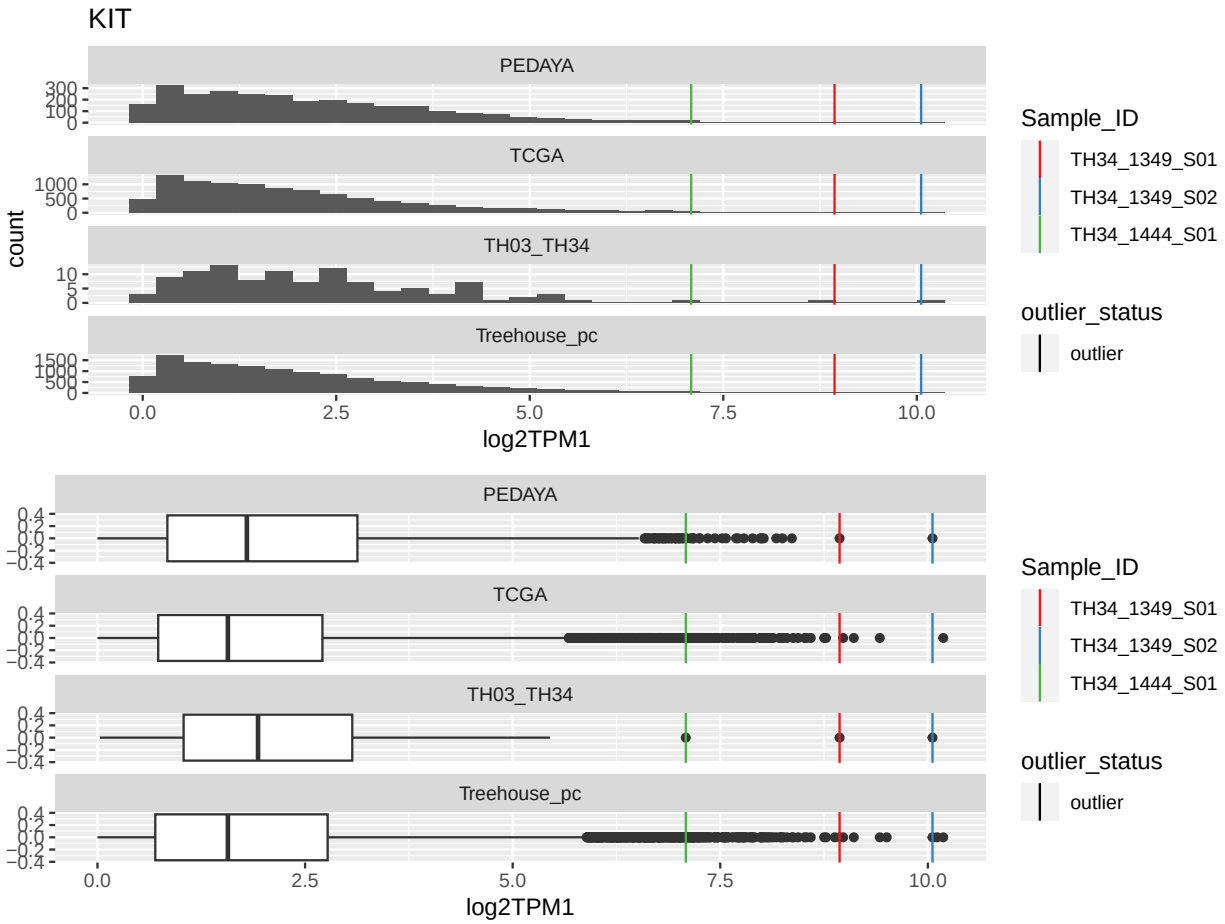
```
##
## [[29]]
```



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

cohort	q25	median	q75	IQR	up_outlier_threshold
PEDAYA	0.06	2.03	3.22	3.17	7.98
TCGA	1.69	2.50	3.41	1.72	6.00
TH03_TH34	2.48	3.28	4.10	1.62	6.53
Treehouse_pc	1.36	2.36	3.33	1.97	6.28

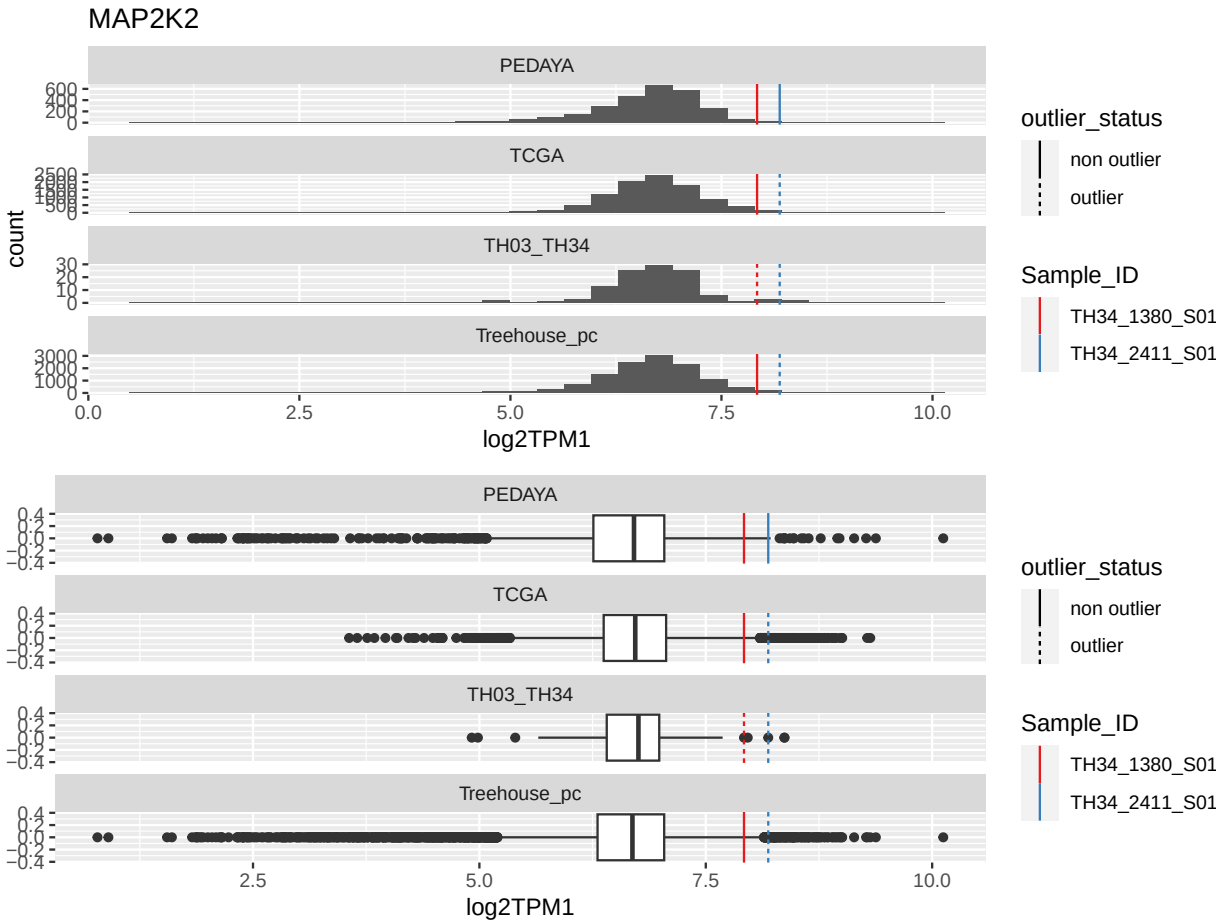
```
##
## [[30]]
```



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

cohort	q25	median	q75	IQR	up_outlier_threshold
PEDAYA	0.84	1.80	3.13	2.29	6.56
TCGA	0.73	1.57	2.71	1.98	5.68
TH03_TH34	1.04	1.93	3.07	2.03	6.11
Treehouse_pc	0.70	1.57	2.77	2.08	5.89

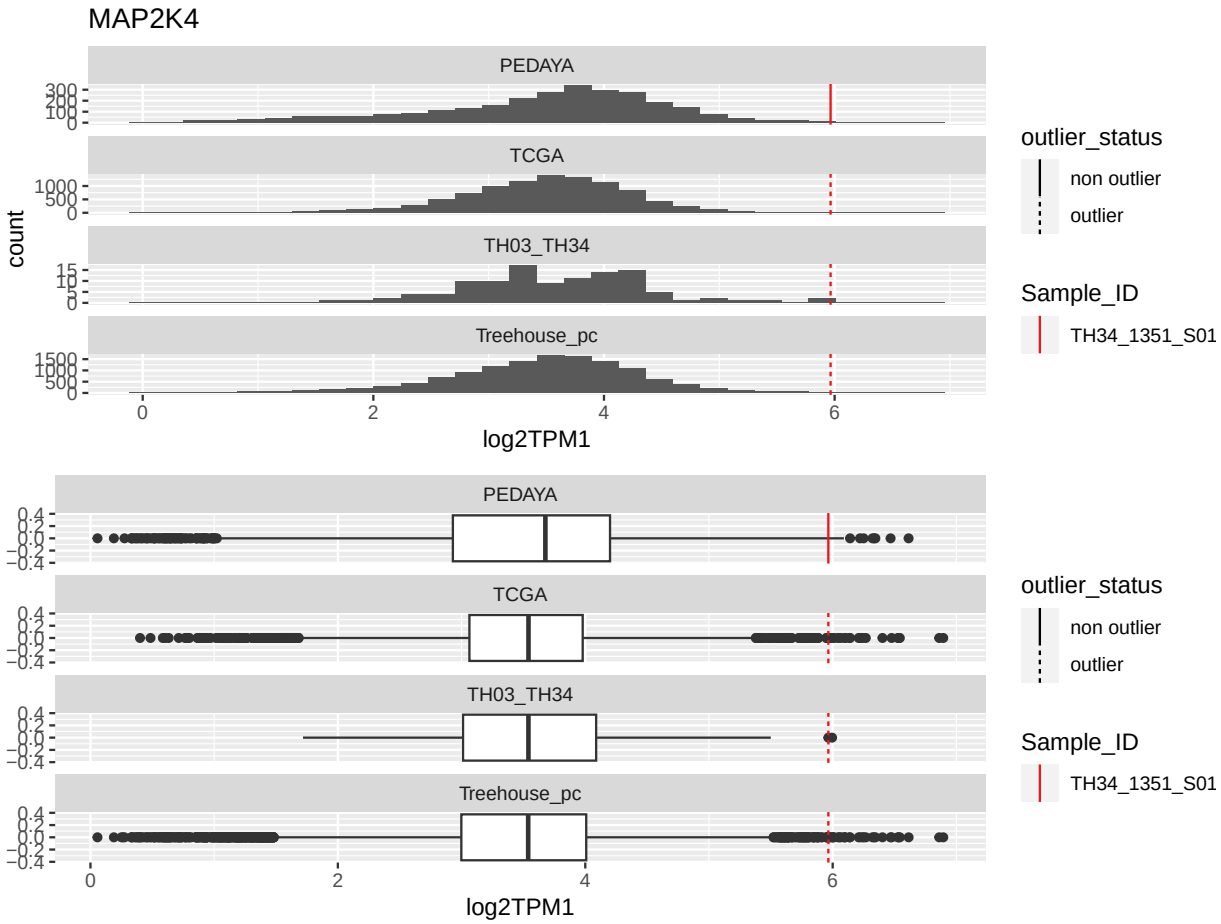
```
##
## [[31]]
```



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

cohort	q25	median	q75	IQR	up_outlier_threshold
PEDAYA	6.26	6.71	7.04	0.78	8.22
TCGA	6.37	6.72	7.06	0.69	8.10
TH03_TH34	6.41	6.76	6.99	0.58	7.86
Treehouse_pc	6.30	6.69	7.04	0.74	8.14

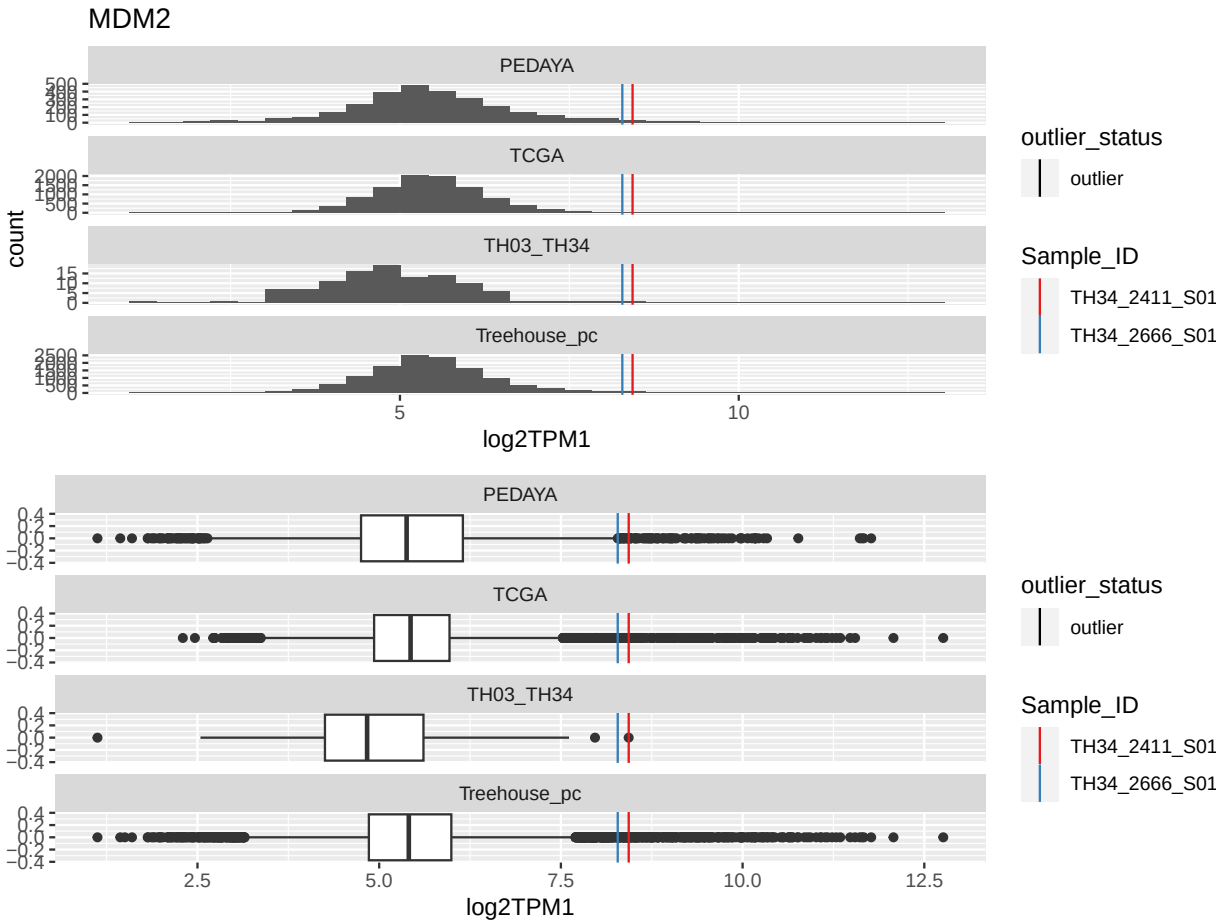
```
##
## [[32]]
```



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

cohort	q25	median	q75	IQR	up_outlier_threshold
PEDAYA	2.93	3.68	4.20	1.27	6.11
TCGA	3.06	3.54	3.98	0.92	5.35
TH03_TH34	3.01	3.54	4.09	1.08	5.70
Treehouse_pc	3.00	3.54	4.01	1.01	5.52

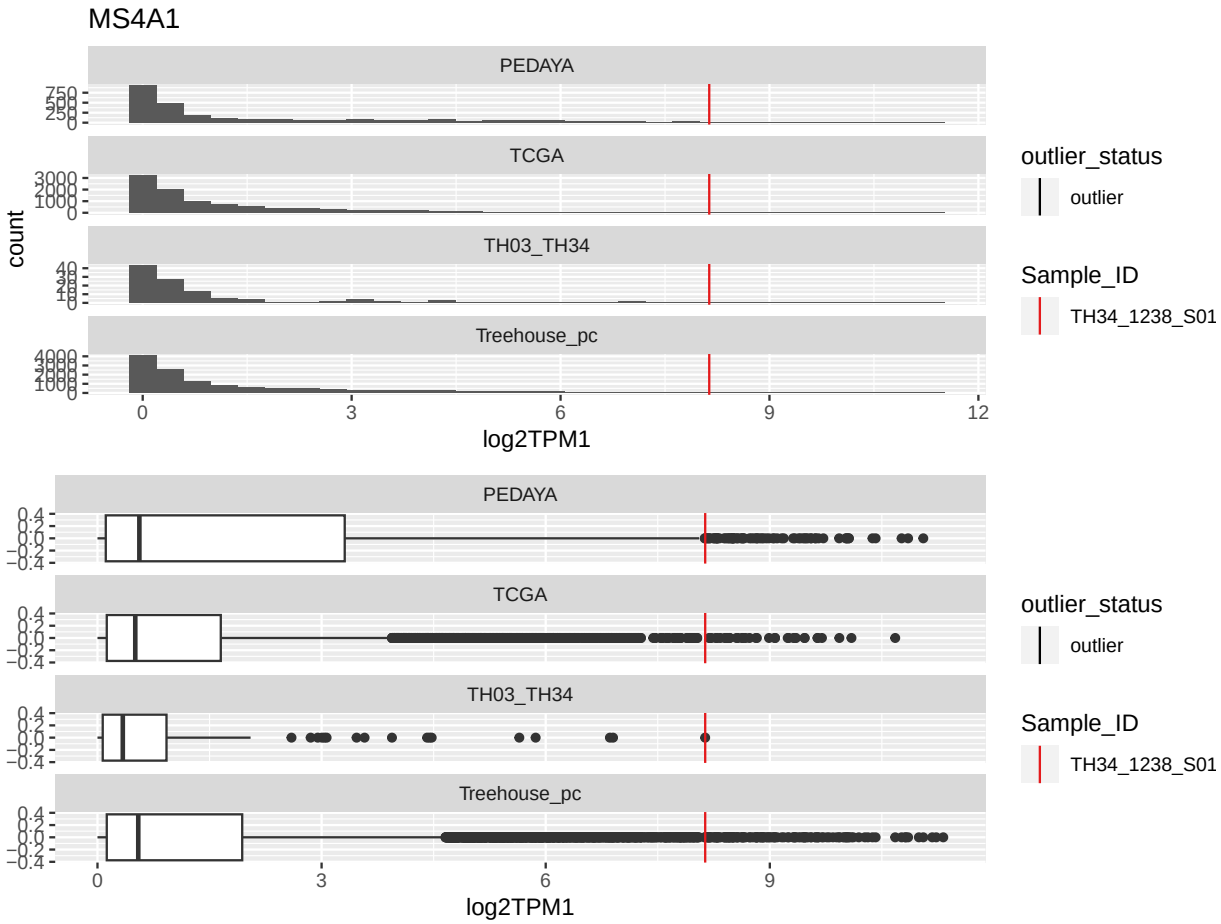
```
##
## [[33]]
```



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

cohort	q25	median	q75	IQR	up_outlier_threshold
PEDAYA	4.75	5.38	6.15	1.40	8.25
TCGA	4.93	5.43	5.97	1.04	7.52
TH03_TH34	4.25	4.83	5.61	1.36	7.64
Treehouse_pc	4.86	5.41	5.99	1.13	7.69

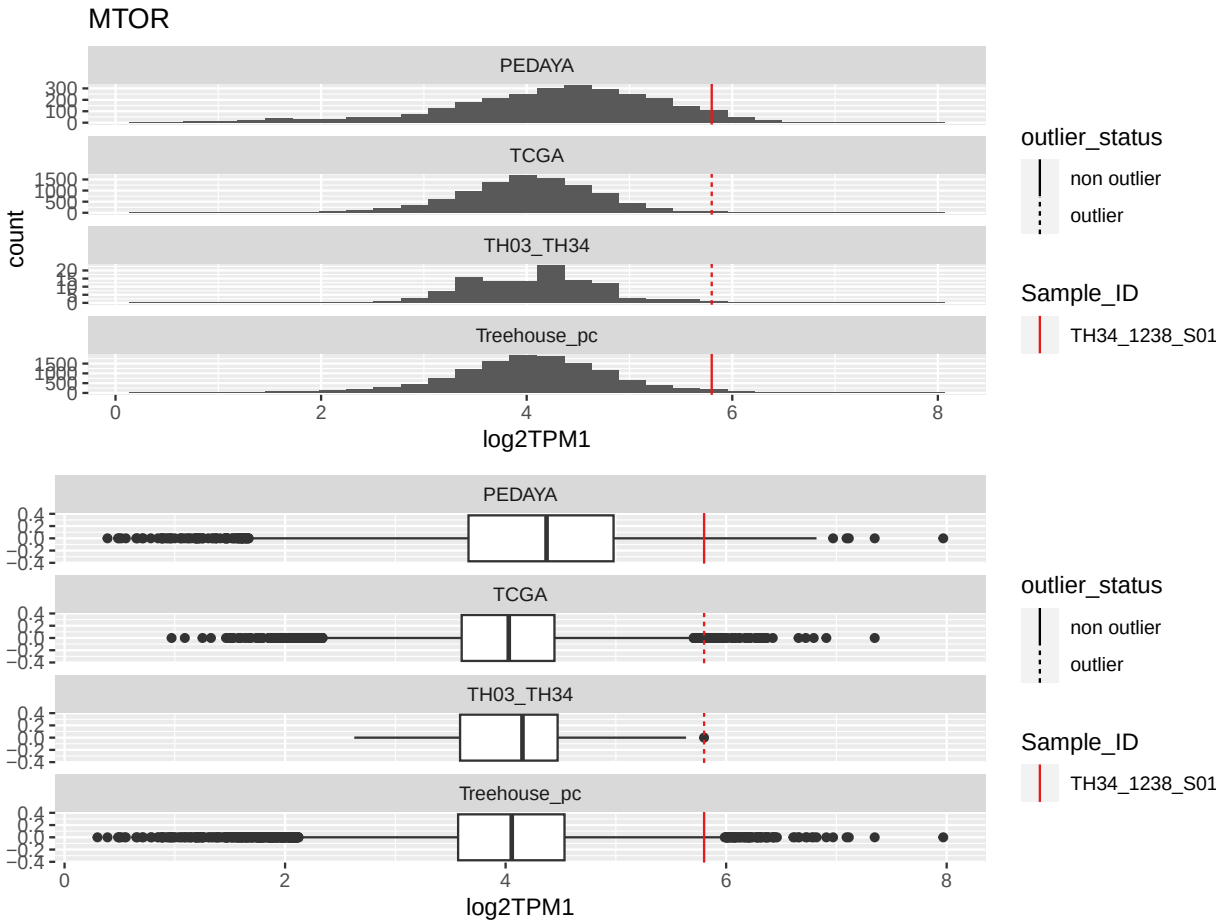
```
##
## [[34]]
```



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

cohort	q25	median	q75	IQR	up_outlier_threshold
PEDAYA	0.11	0.56	3.31	3.20	8.11
TCGA	0.12	0.51	1.65	1.53	3.94
TH03_TH34	0.07	0.34	0.92	0.85	2.20
Treehouse_pc	0.12	0.55	1.94	1.81	4.66

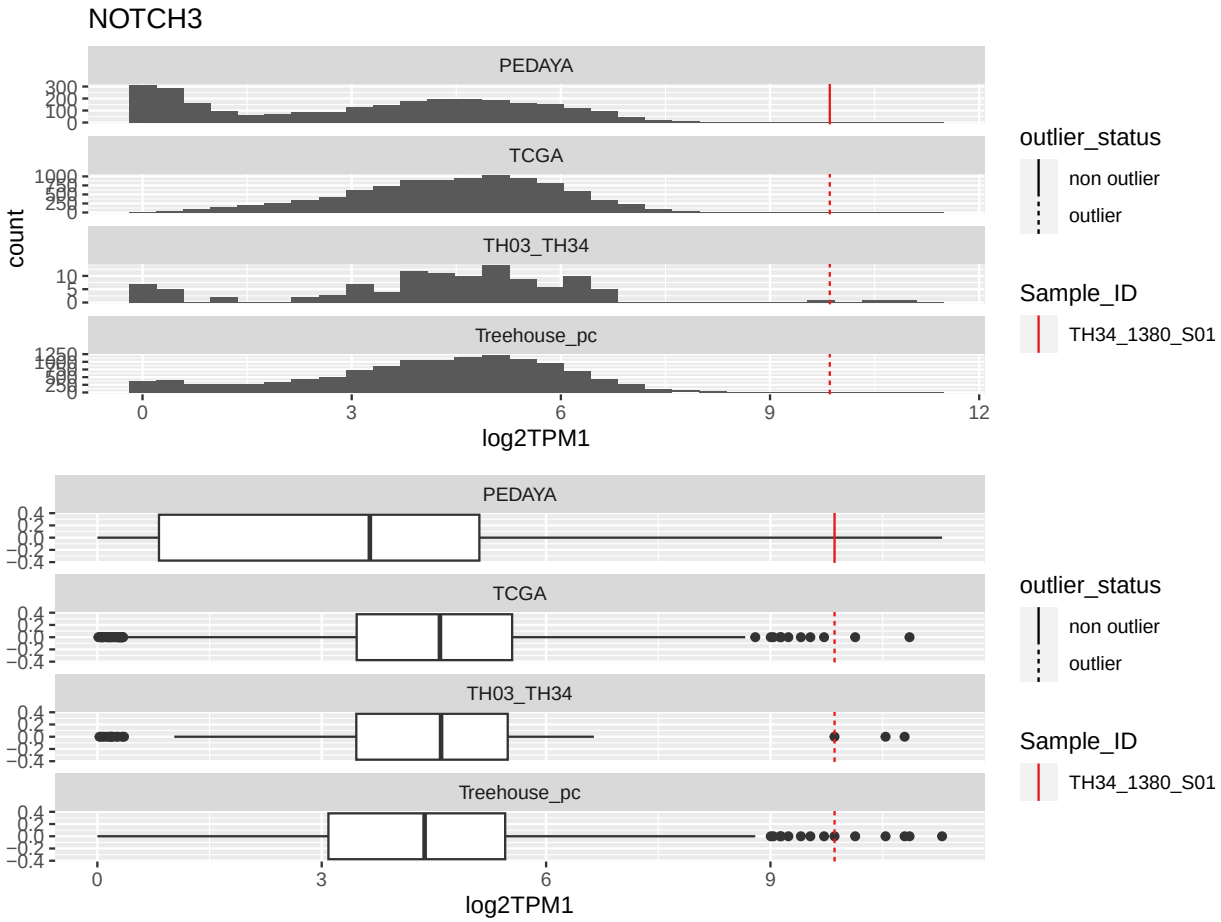
```
##
## [[35]]
```



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

cohort	q25	median	q75	IQR	up_outlier_threshold
PEDAYA	3.66	4.37	4.98	1.32	6.95
TCGA	3.60	4.03	4.44	0.84	5.70
TH03_TH34	3.59	4.15	4.47	0.88	5.80
Treehouse_pc	3.57	4.06	4.53	0.96	5.98

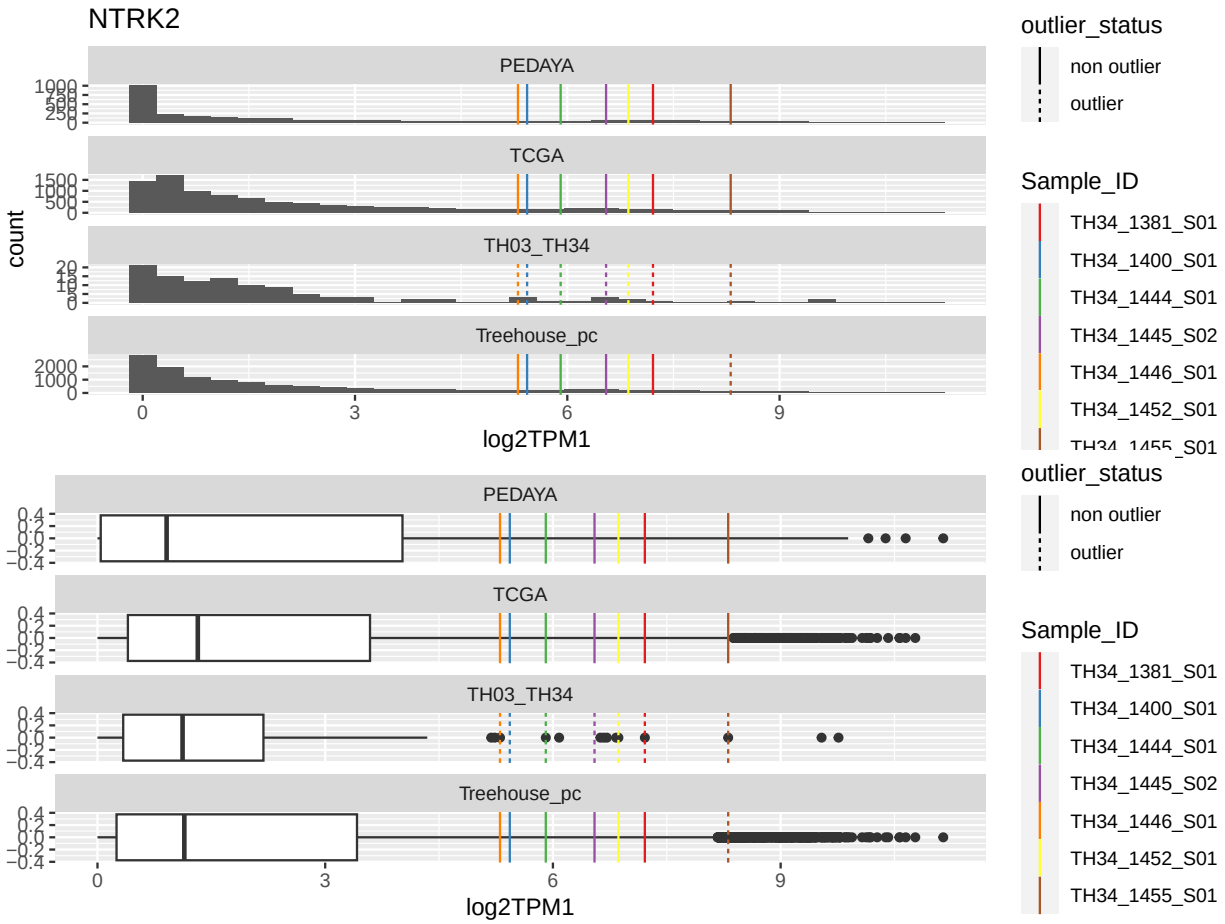
```
##
## [[36]]
```



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

cohort	q25	median	q75	IQR	up_outlier_threshold
PEDAYA	0.82	3.64	5.11	4.28	11.53
TCGA	3.47	4.58	5.55	2.08	8.66
TH03_TH34	3.46	4.60	5.49	2.02	8.52
Treehouse_pc	3.09	4.38	5.45	2.36	9.00

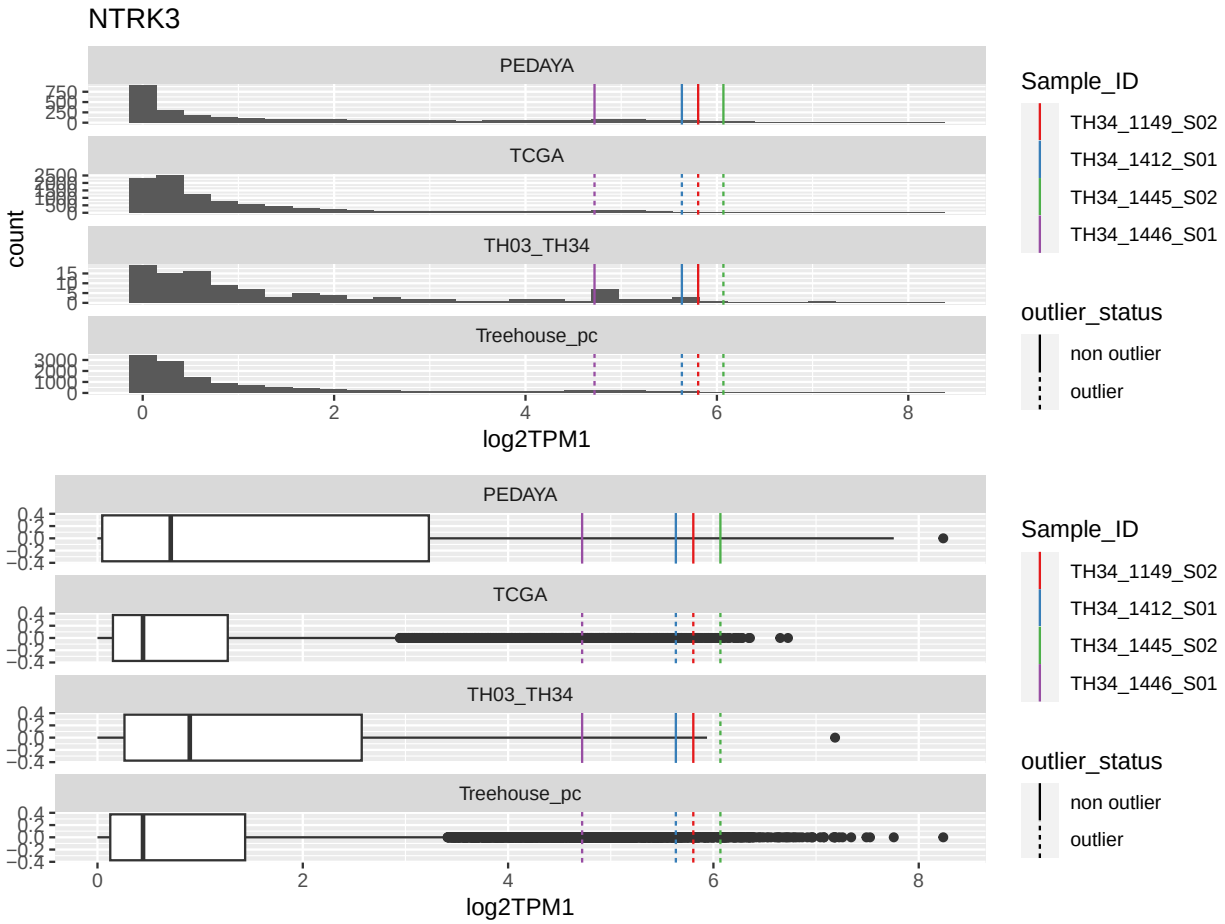
```
##
## [[37]]
```



Sample_ID	gene	PEDAYA	TCGA	TH03_TH34	Treehouse_pc	log2TPM1
TH34_1400_S01	NTRK2	FALSE	FALSE	TRUE	FALSE	5.432
TH34_1444_S01	NTRK2	FALSE	FALSE	TRUE	FALSE	5.906
TH34_1445_S02	NTRK2	FALSE	FALSE	TRUE	FALSE	6.548
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

cohort	q25	median	q75	IQR	up_outlier_threshold
PEDAYA	0.04	0.91	4.02	3.98	9.98
TCGA	0.40	1.32	3.59	3.19	8.37
TH03_TH34	0.34	1.12	2.19	1.85	4.96
Treehouse_pc	0.25	1.14	3.42	3.17	8.17

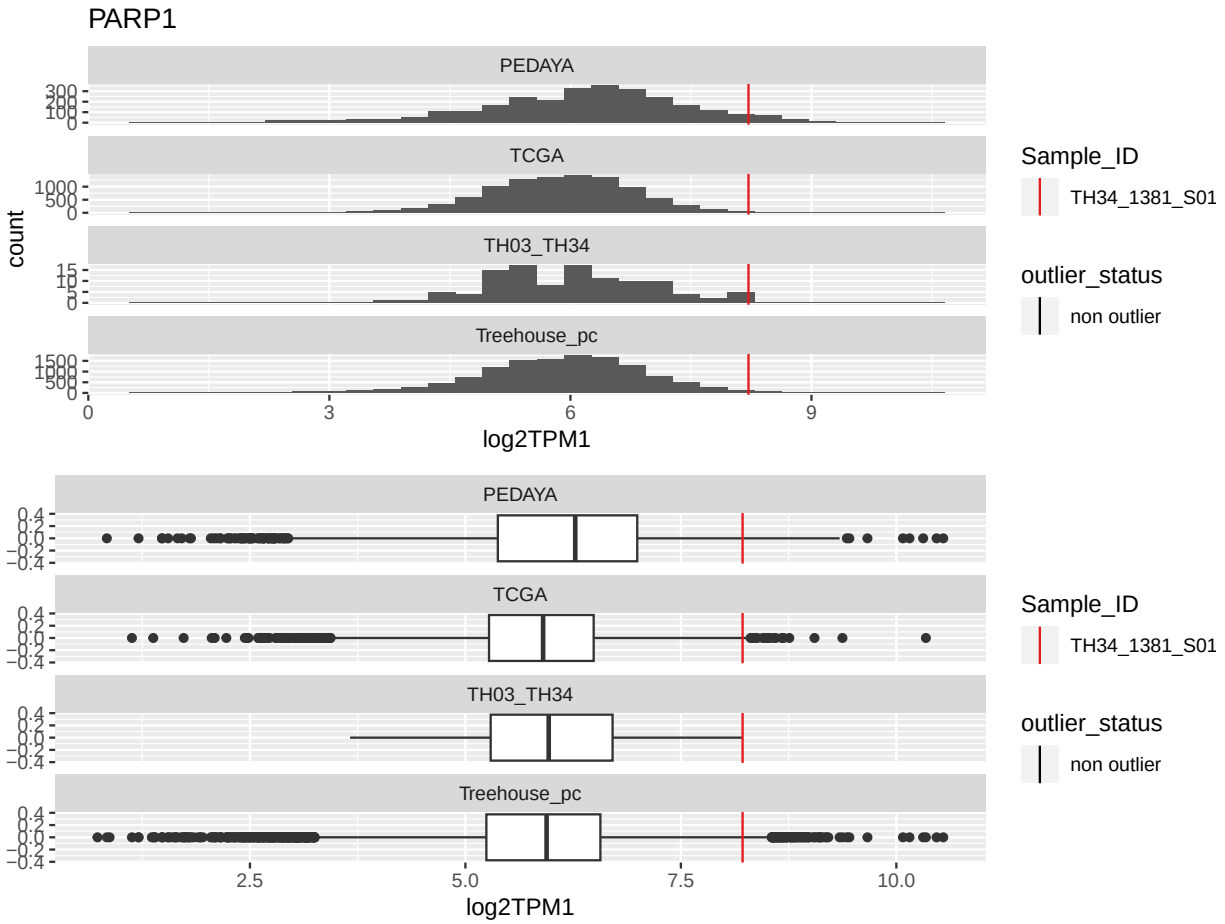
```
##
## [[38]]
```



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_1446_S01	NTRK3	FALSE	TRUE	FALSE	TRUE	4.721

cohort	q25	median	q75	IQR	up_outlier_threshold
PEDAYA	0.05	0.71	3.23	3.18	8.00
TCGA	0.15	0.44	1.27	1.12	2.95
TH03_TH34	0.26	0.90	2.57	2.31	6.04
Treehouse_pc	0.12	0.44	1.44	1.31	3.41

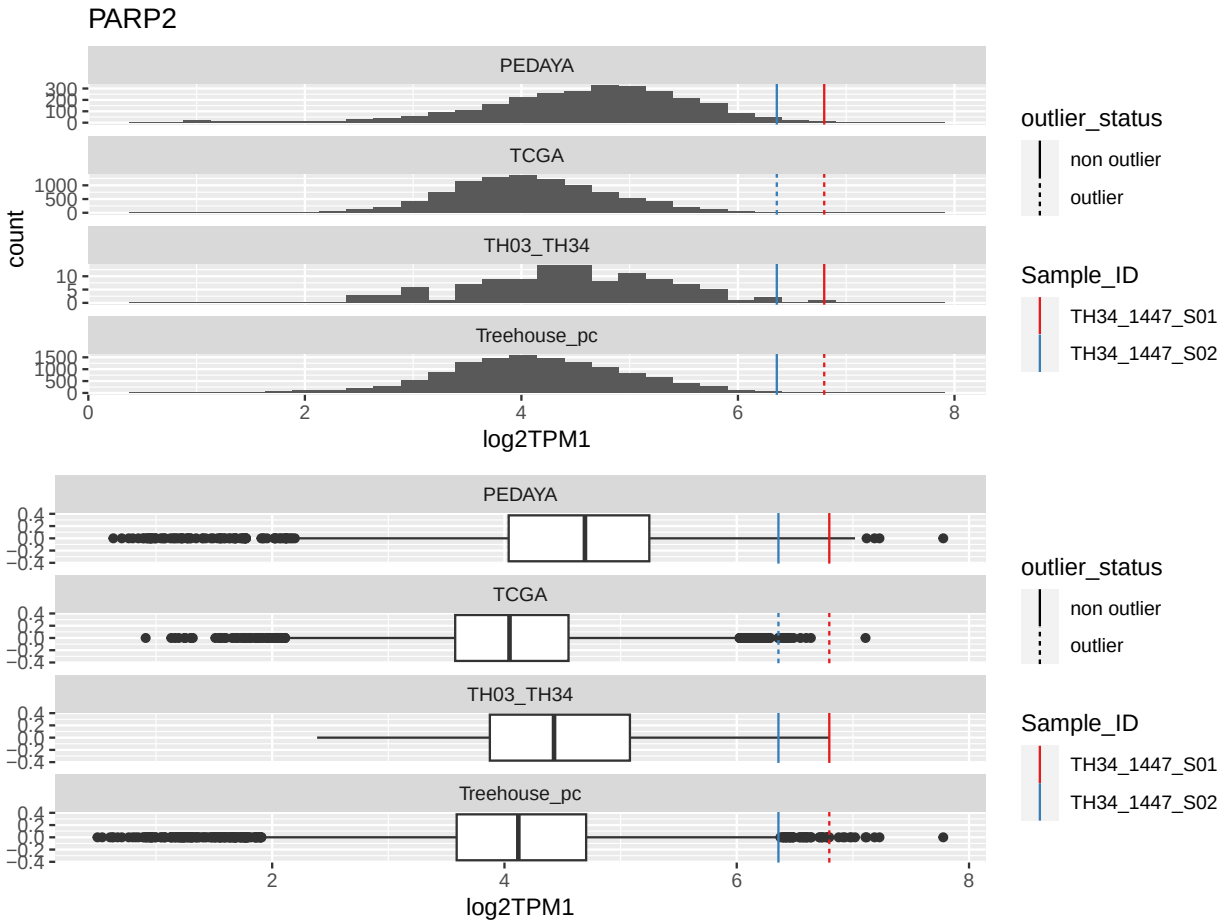
```
##
## [[39]]
```



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

cohort	q25	median	q75	IQR	up_outlier_threshold
PEDAYA	5.38	6.27	6.99	1.62	9.42
TCGA	5.27	5.90	6.49	1.21	8.31
TH03_TH34	5.29	5.97	6.71	1.41	8.83
Treehouse_pc	5.24	5.94	6.57	1.32	8.55

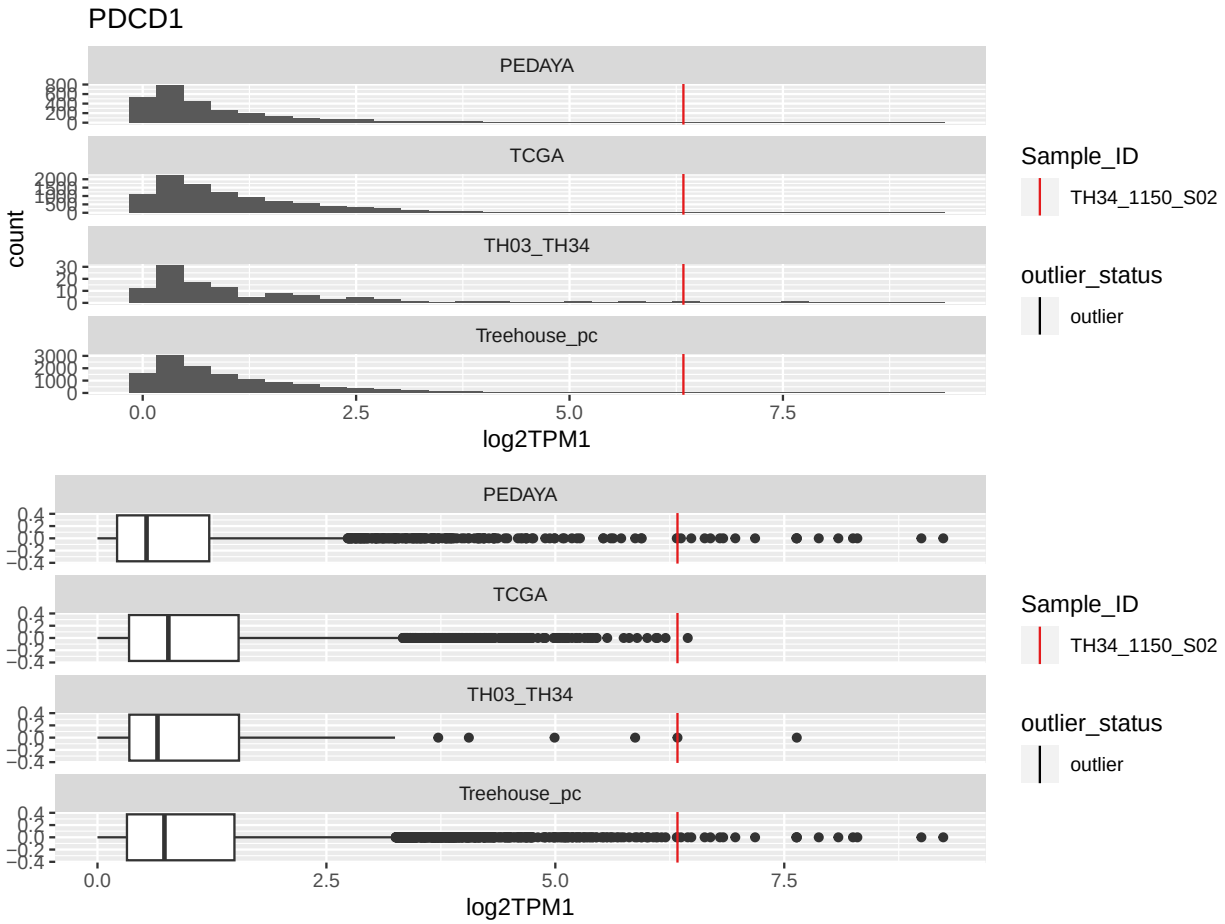
```
##
## [[40]]
```



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

cohort	q25	median	q75	IQR	up_outlier_threshold
PEDAYA	4.04	4.69	5.25	1.21	7.06
TCGA	3.58	4.04	4.55	0.98	6.02
TH03_TH34	3.87	4.43	5.08	1.21	6.89
Treehouse_pc	3.59	4.12	4.70	1.11	6.38

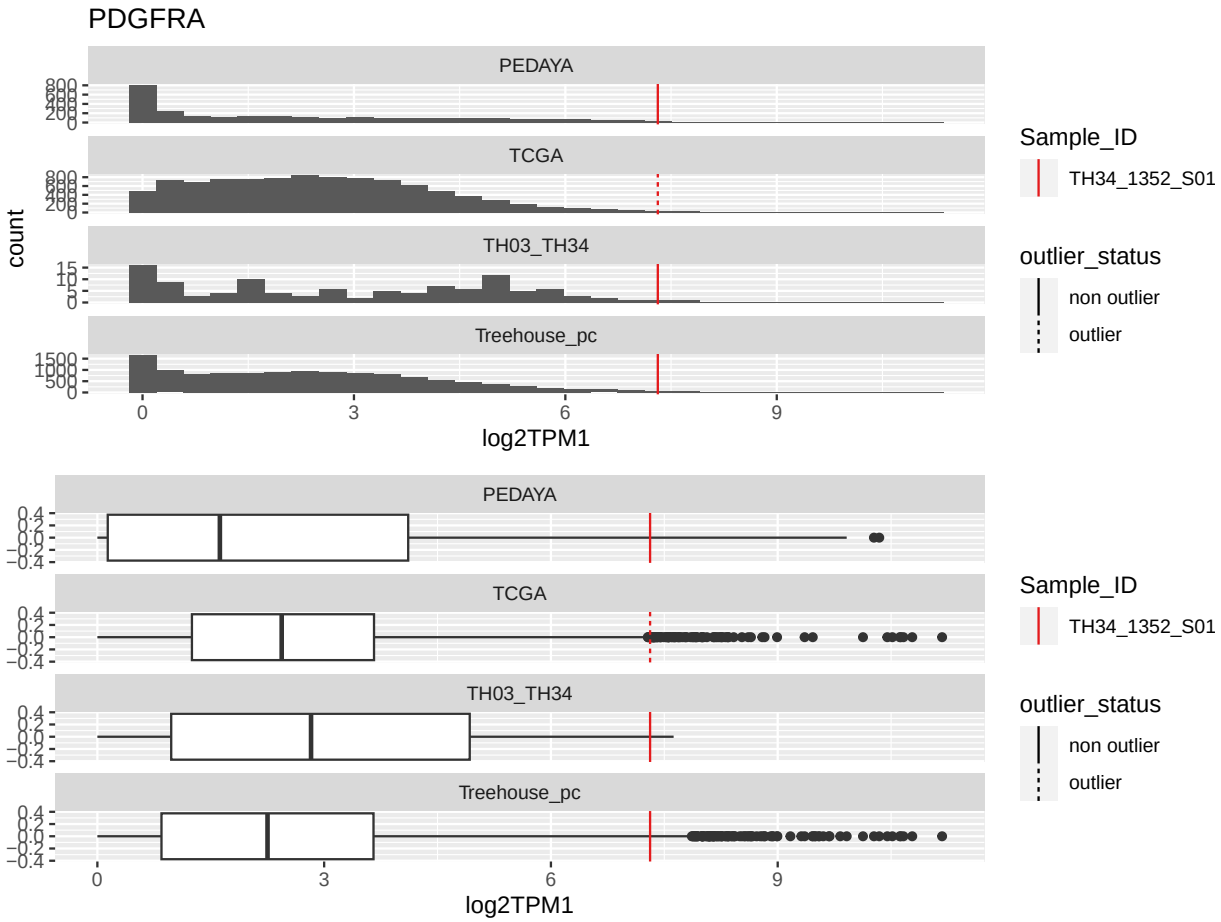
```
##
## [[41]]
```



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

cohort	q25	median	q75	IQR	up_outlier_threshold
PEDAYA	0.21	0.54	1.22	1.01	2.73
TCGA	0.34	0.77	1.54	1.20	3.34
TH03_TH34	0.35	0.66	1.55	1.20	3.34
Treehouse_pc	0.32	0.73	1.50	1.17	3.26

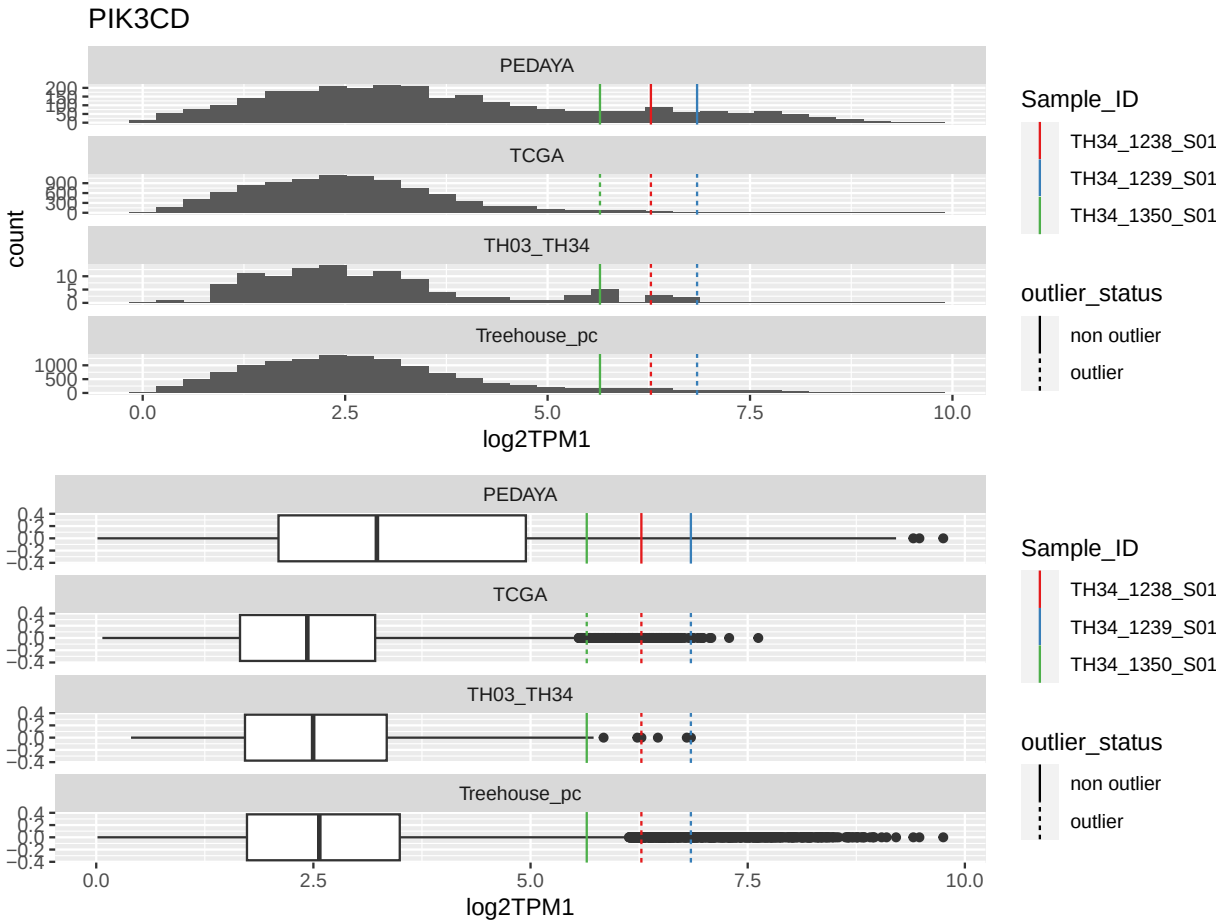
```
##
## [[42]]
```



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

cohort	q25	median	q75	IQR	up_outlier_threshold
PEDAYA	0.14	1.62	4.11	3.97	10.07
TCGA	1.25	2.44	3.66	2.41	7.27
TH03_TH34	0.98	2.83	4.93	3.95	10.85
Treehouse_pc	0.85	2.25	3.65	2.80	7.86

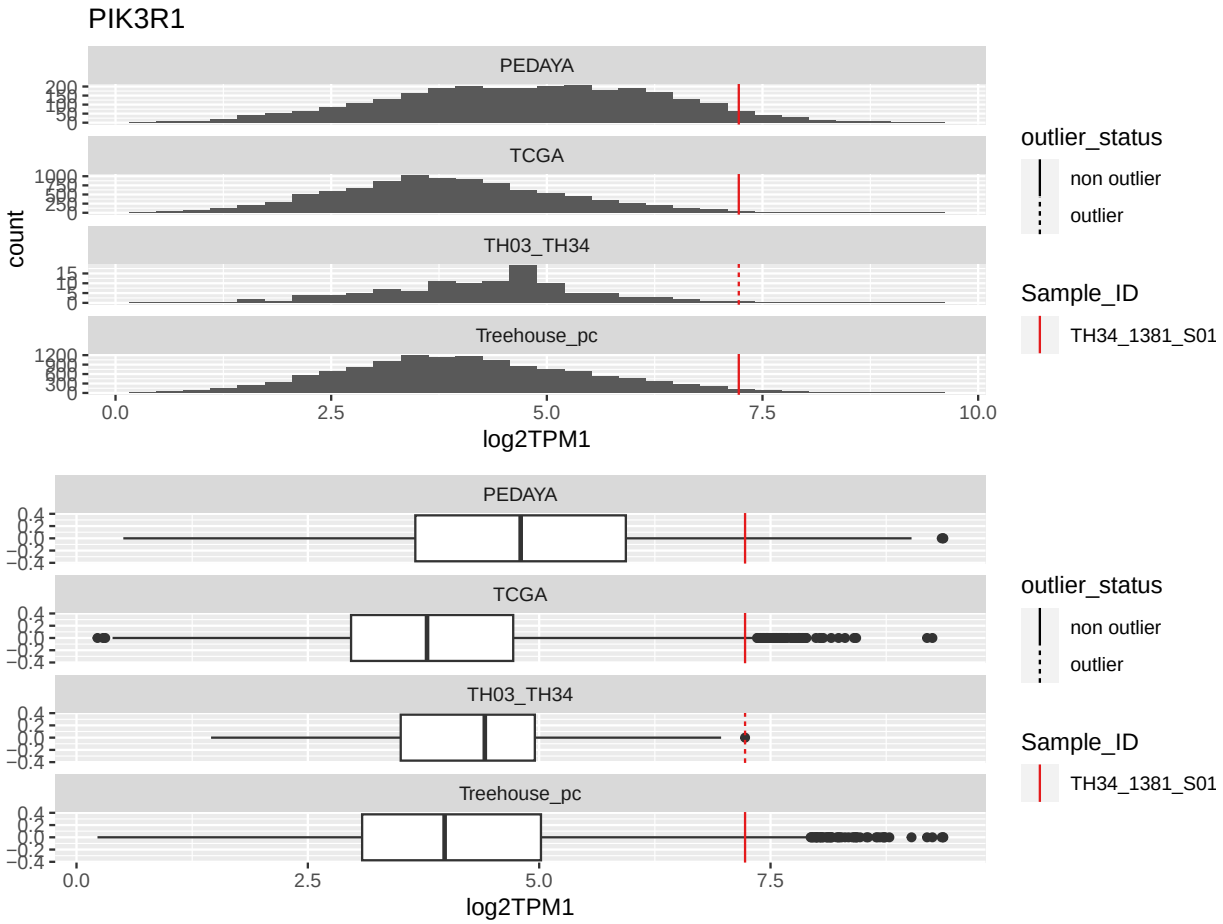
```
##
## [[43]]
```



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

cohort	q25	median	q75	IQR	up_outlier_threshold
PEDAYA	2.10	3.23	4.95	2.85	9.22
TCGA	1.66	2.43	3.21	1.56	5.54
TH03_TH34	1.71	2.50	3.34	1.63	5.79
Treehouse_pc	1.74	2.57	3.49	1.76	6.13

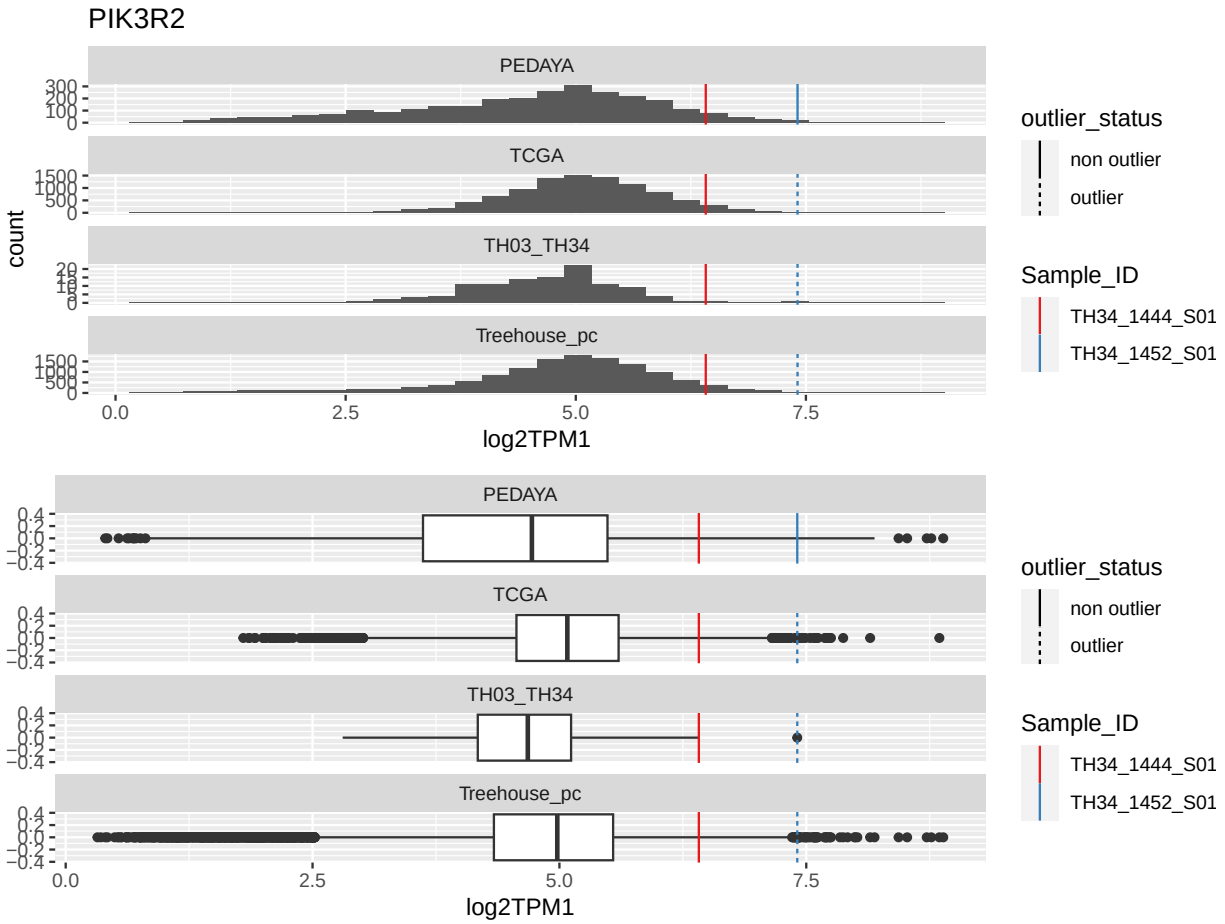
```
##
## [[44]]
```



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

cohort	q25	median	q75	IQR	up_outlier_threshold
PEDAYA	3.66	4.80	5.94	2.28	9.35
TCGA	2.97	3.79	4.72	1.75	7.35
TH03_TH34	3.50	4.41	4.95	1.45	7.13
Treehouse_pc	3.09	3.98	5.02	1.93	7.92

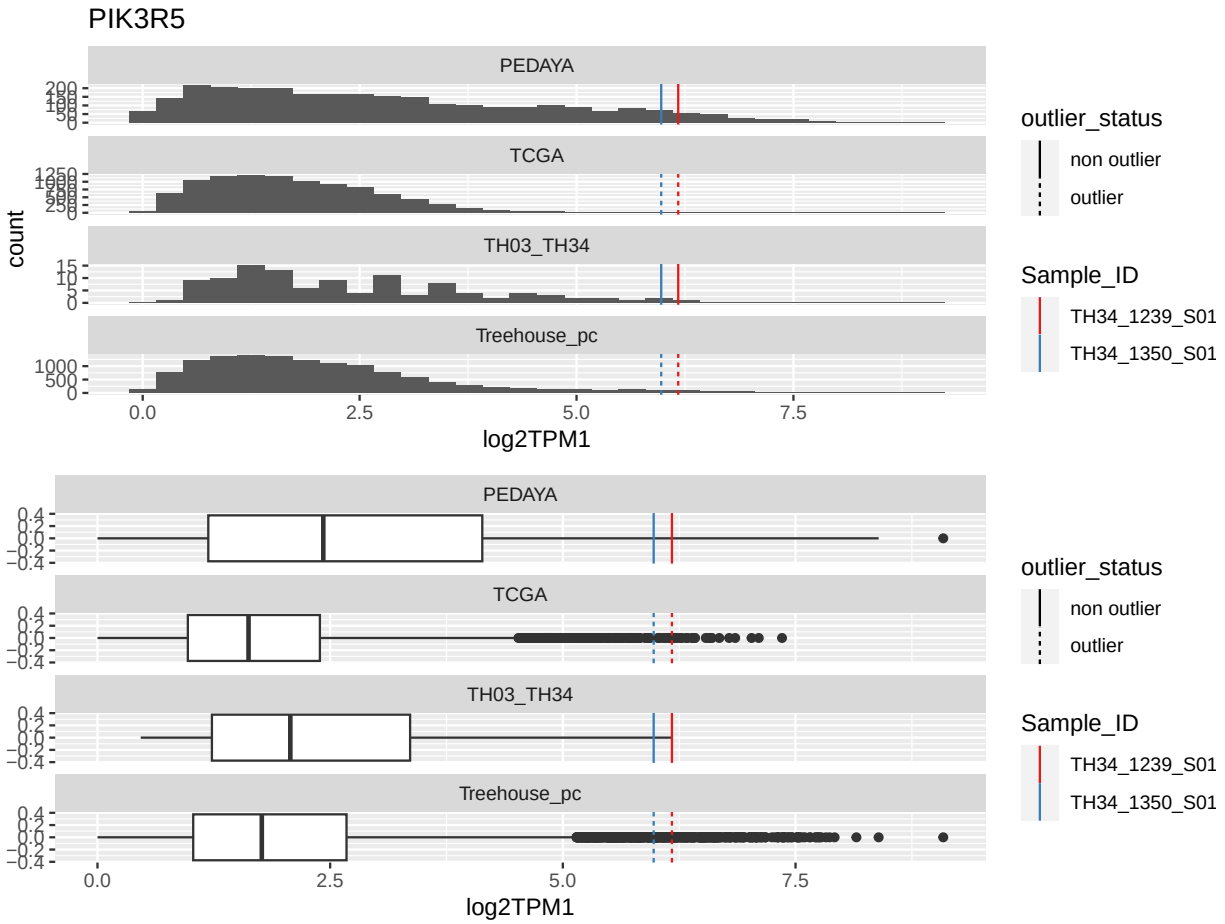
```
##
## [[45]]
```



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

cohort	q25	median	q75	IQR	up_outlier_threshold
PEDAYA	3.62	4.72	5.49	1.87	8.29
TCGA	4.56	5.08	5.60	1.03	7.15
TH03_TH34	4.17	4.68	5.12	0.94	6.54
Treehouse_pc	4.34	4.98	5.54	1.21	7.35

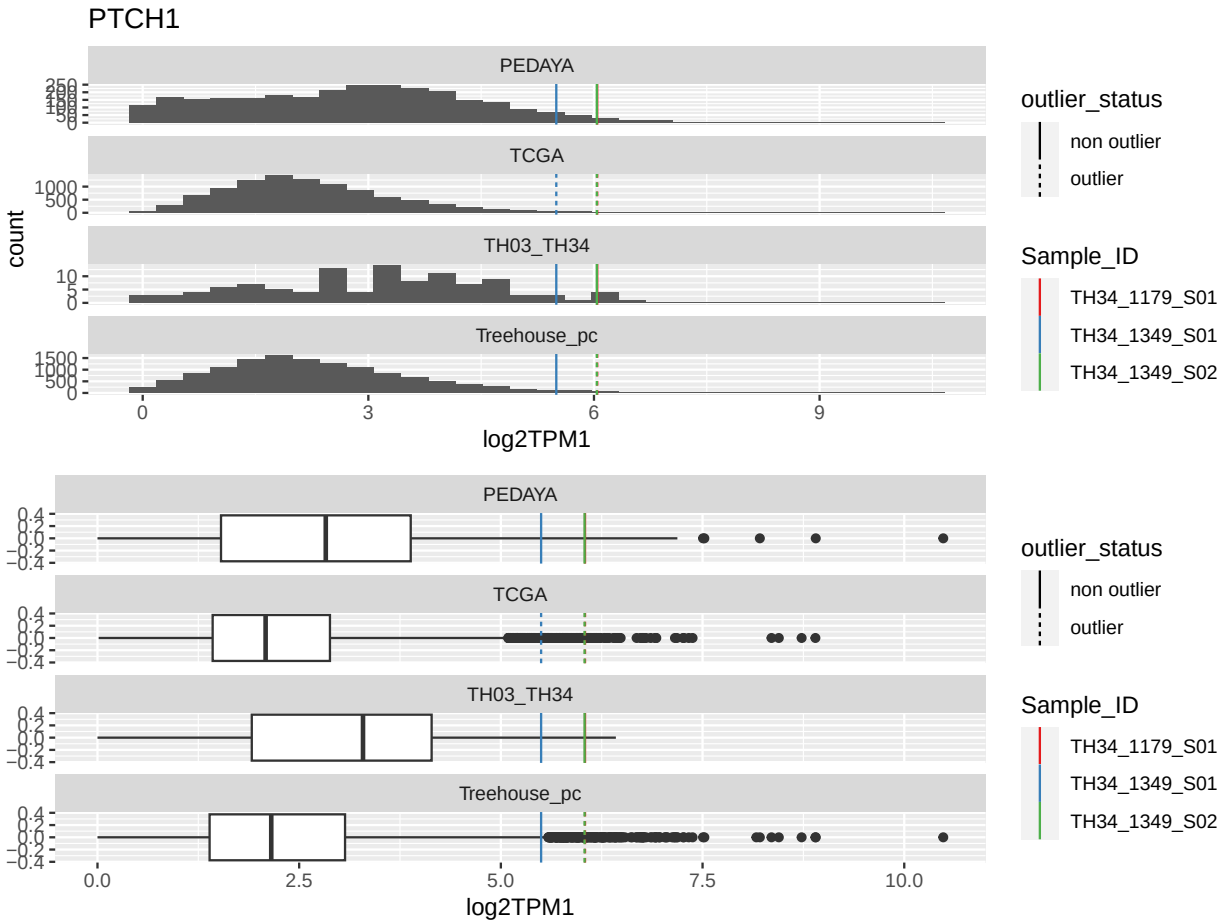
```
##
## [[46]]
```



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

cohort	q25	median	q75	IQR	up_outlier_threshold
PEDAYA	1.19	2.43	4.13	2.94	8.55
TCGA	0.97	1.62	2.39	1.42	4.52
TH03_TH34	1.23	2.07	3.36	2.13	6.55
Treehouse_pc	1.03	1.77	2.68	1.65	5.15

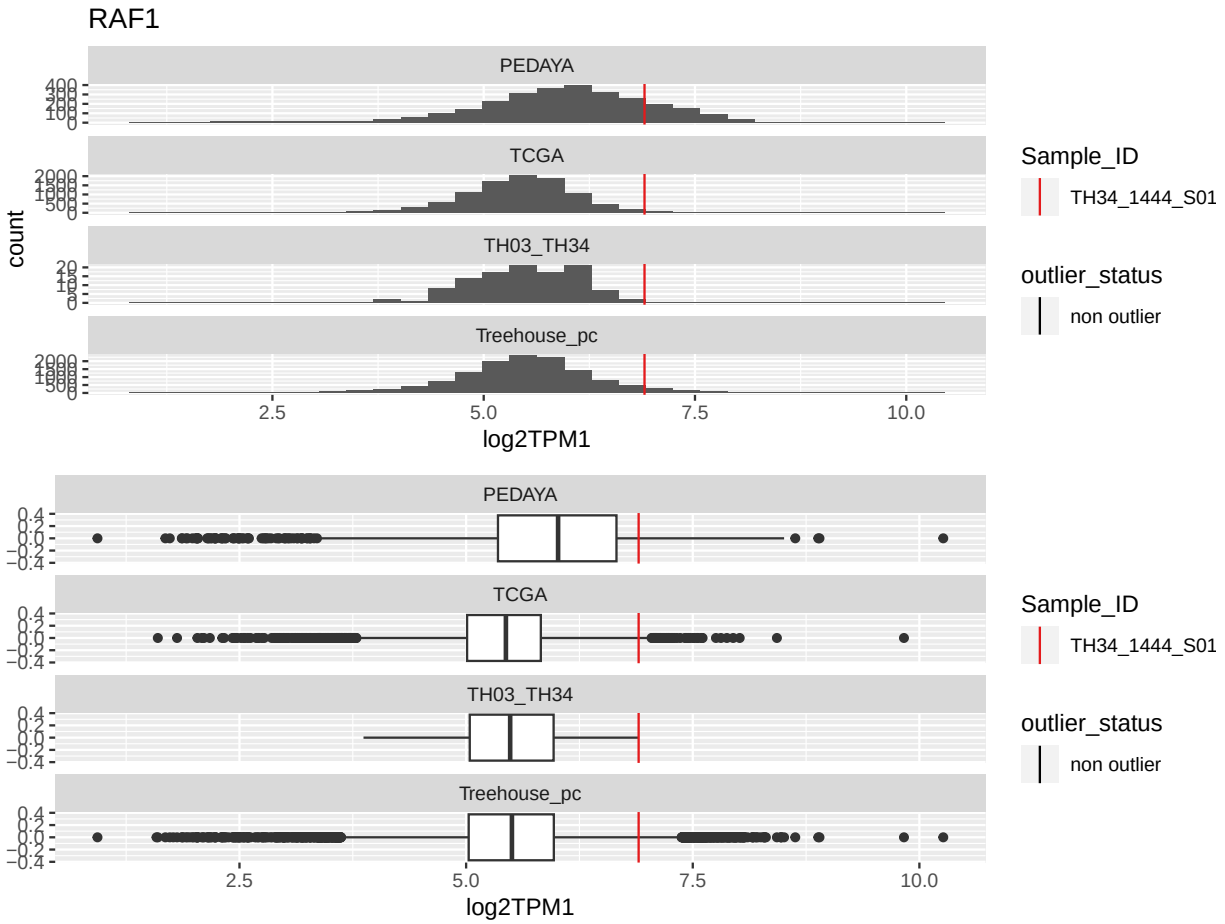
```
##
## [[47]]
```



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

cohort	q25	median	q75	IQR	up_outlier_threshold
PEDAYA	1.53	2.83	3.88	2.35	7.41
TCGA	1.43	2.08	2.88	1.45	5.06
TH03_TH34	1.91	3.29	4.14	2.23	7.49
Treehouse_pc	1.39	2.15	3.07	1.68	5.59

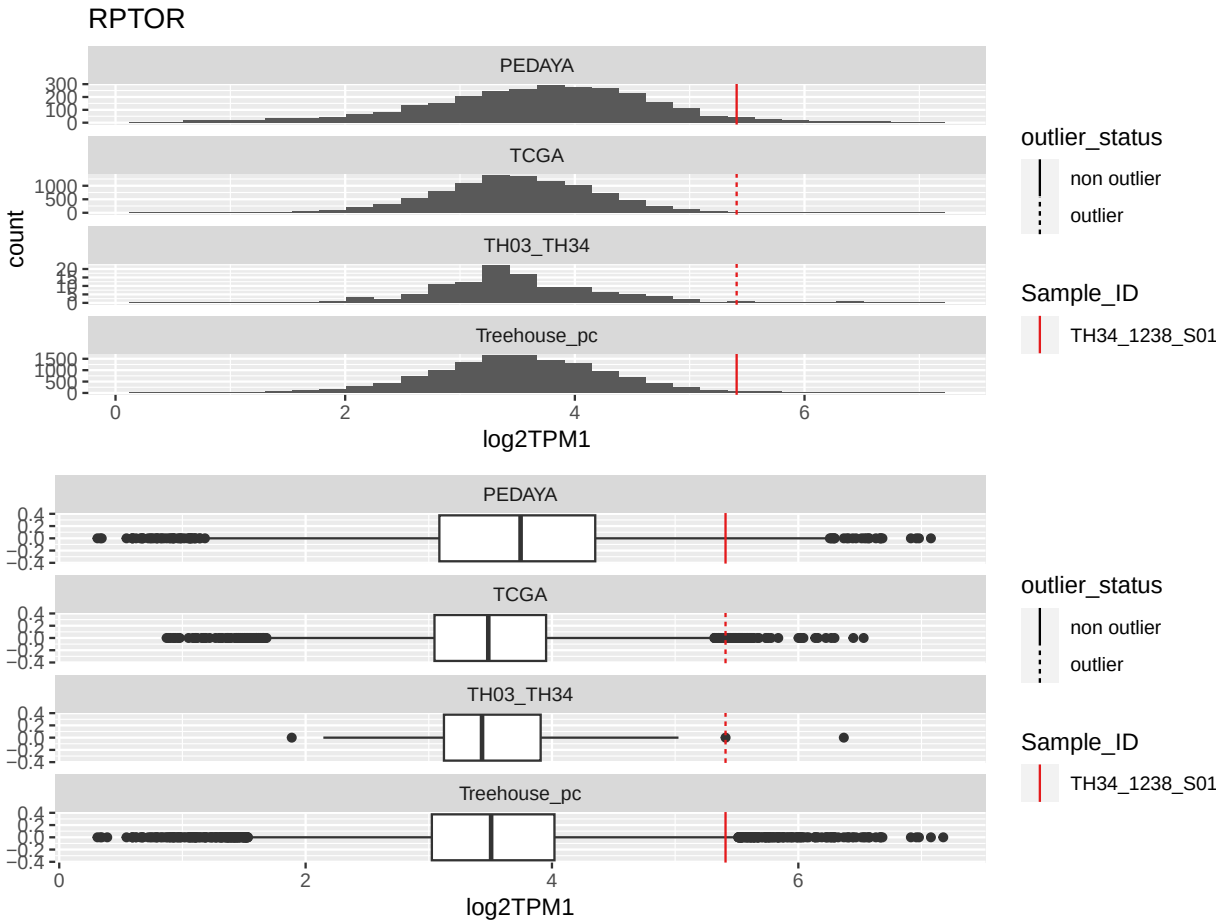
```
##
## [[48]]
```



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

cohort	q25	median	q75	IQR	up_outlier_threshold
PEDAYA	5.35	6.01	6.66	1.31	8.62
TCGA	5.01	5.44	5.82	0.81	7.04
TH03_TH34	5.04	5.49	5.97	0.92	7.35
Treehouse_pc	5.03	5.51	5.97	0.94	7.38

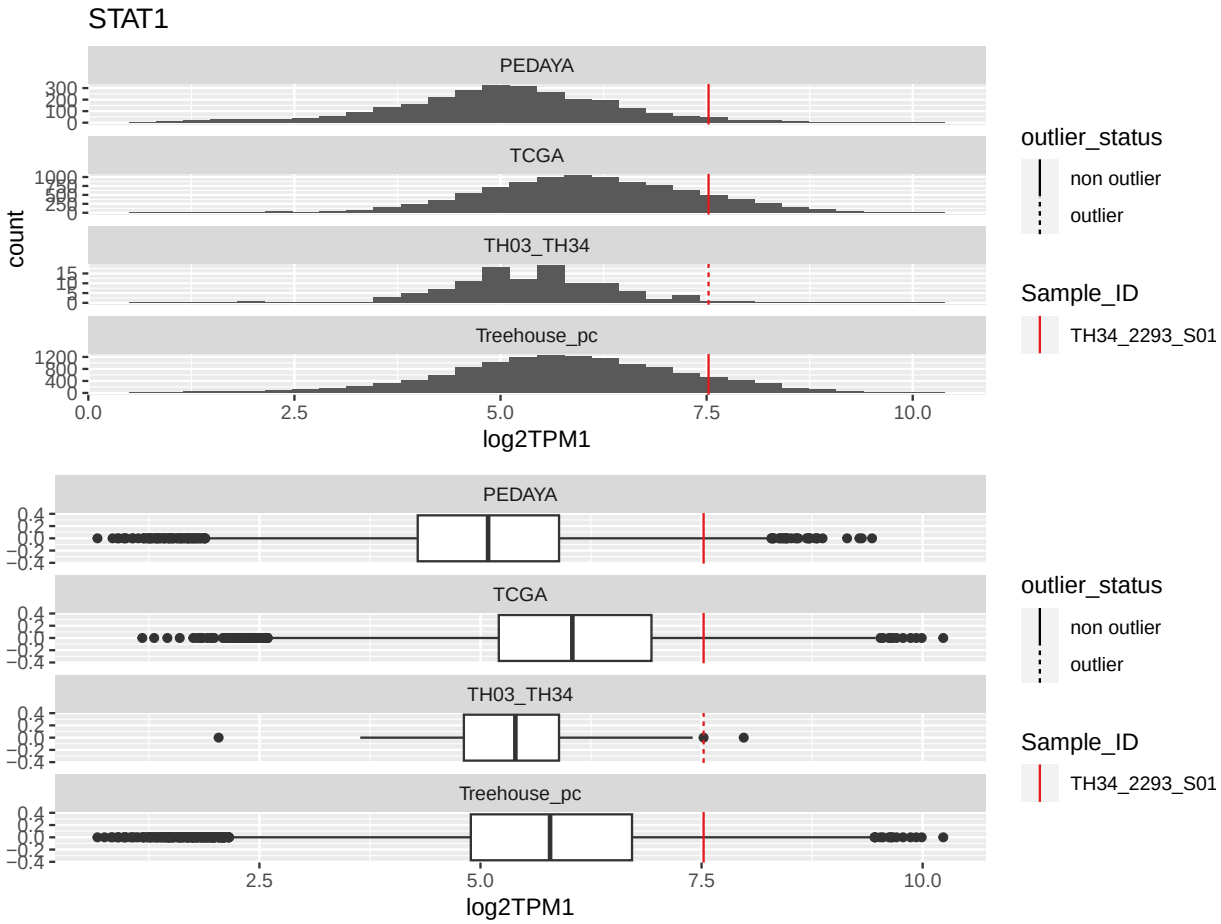
```
##
## [[49]]
```



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

cohort	q25	median	q75	IQR	up_outlier_threshold
PEDAYA	3.09	3.75	4.35	1.27	6.25
TCGA	3.05	3.48	3.95	0.91	5.31
TH03_TH34	3.12	3.43	3.91	0.79	5.09
Treehouse_pc	3.03	3.51	4.02	1.00	5.51

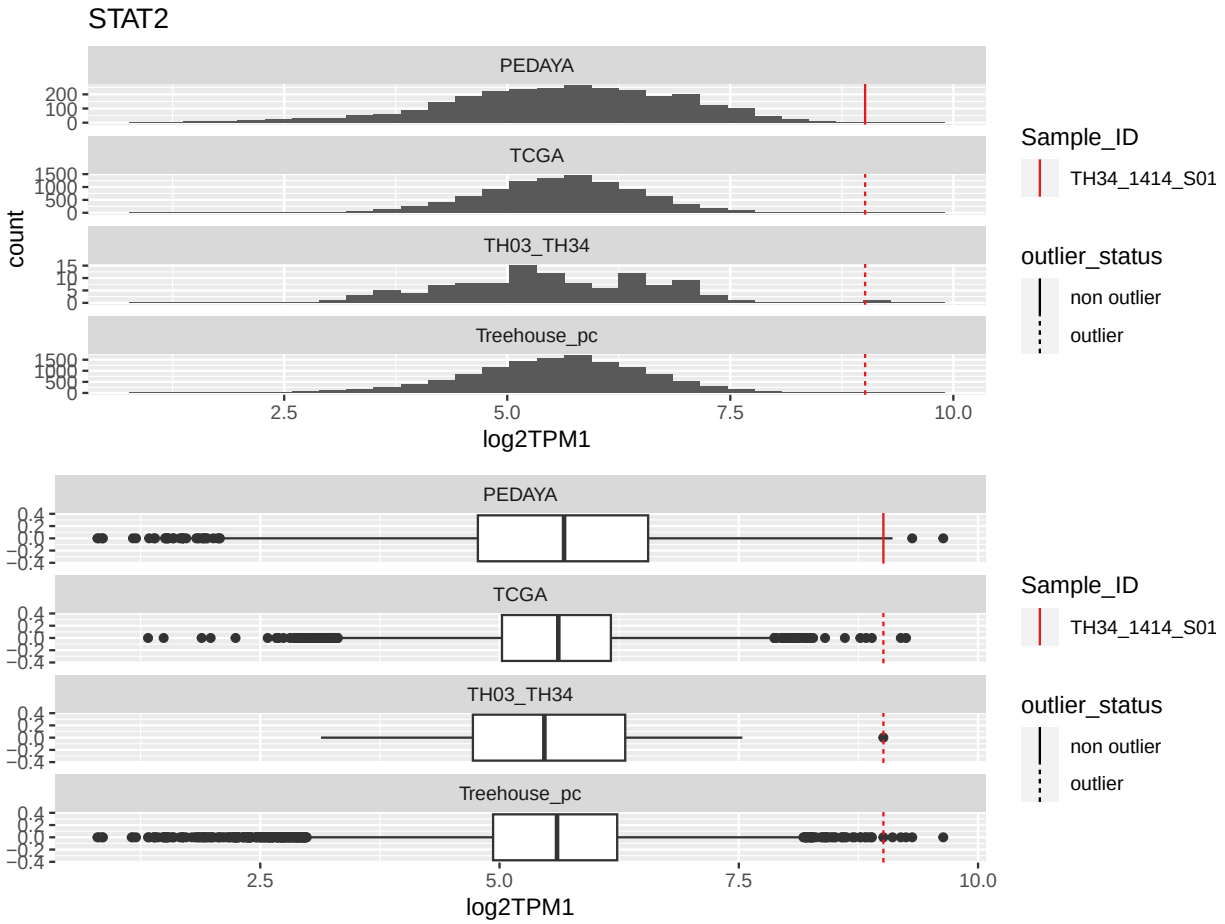
```
##
## [[50]]
```



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

cohort	q25	median	q75	IQR	up_outlier_threshold
PEDAYA	4.29	5.09	5.89	1.60	8.29
TCGA	5.21	6.04	6.93	1.73	9.52
TH03_TH34	4.81	5.39	5.89	1.08	7.50
Treehouse_pc	4.89	5.79	6.71	1.82	9.45

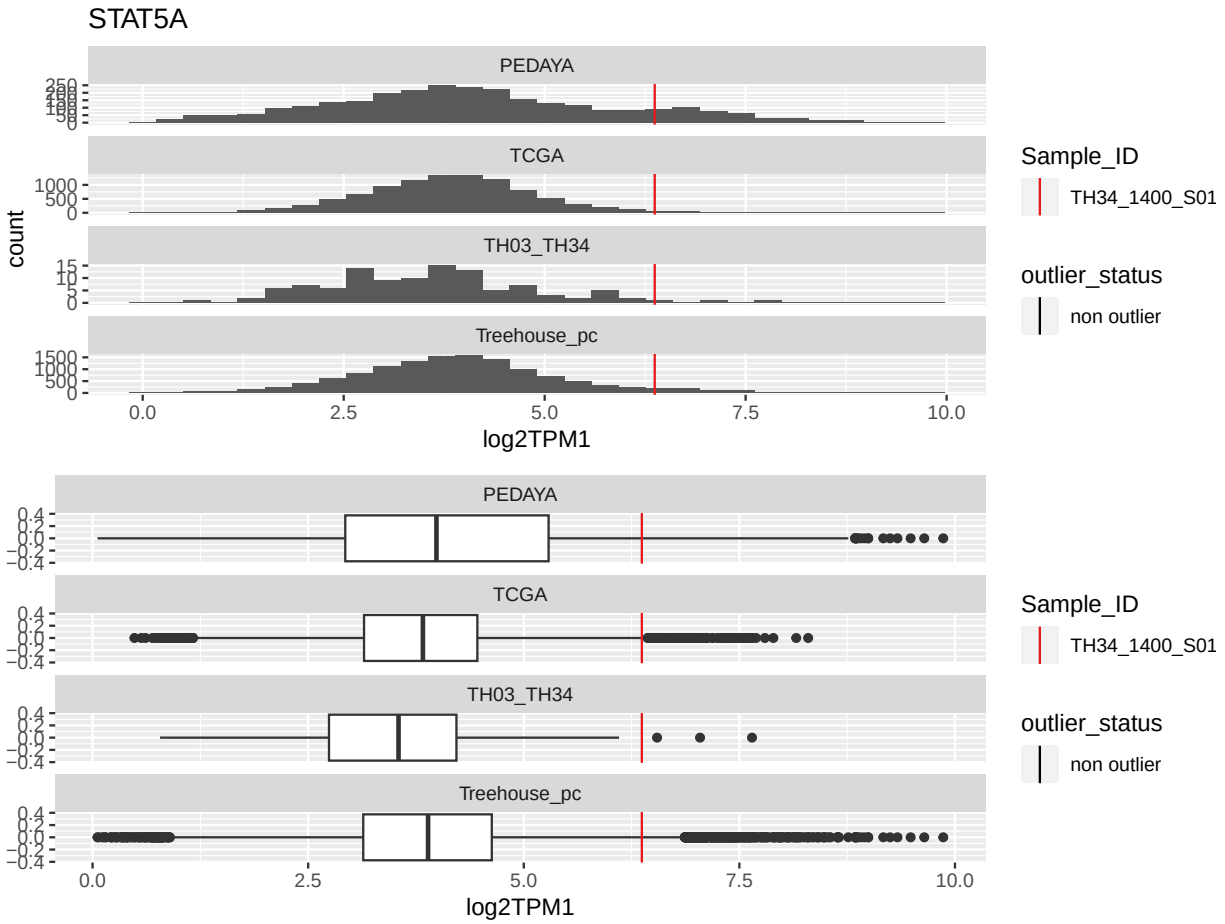
```
##
## [[51]]
```



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

cohort	q25	median	q75	IQR	up_outlier_threshold
PEDAYA	4.77	5.67	6.55	1.78	9.22
TCGA	5.03	5.61	6.16	1.14	7.87
TH03_TH34	4.72	5.47	6.31	1.59	8.70
Treehouse_pc	4.93	5.60	6.23	1.30	8.17

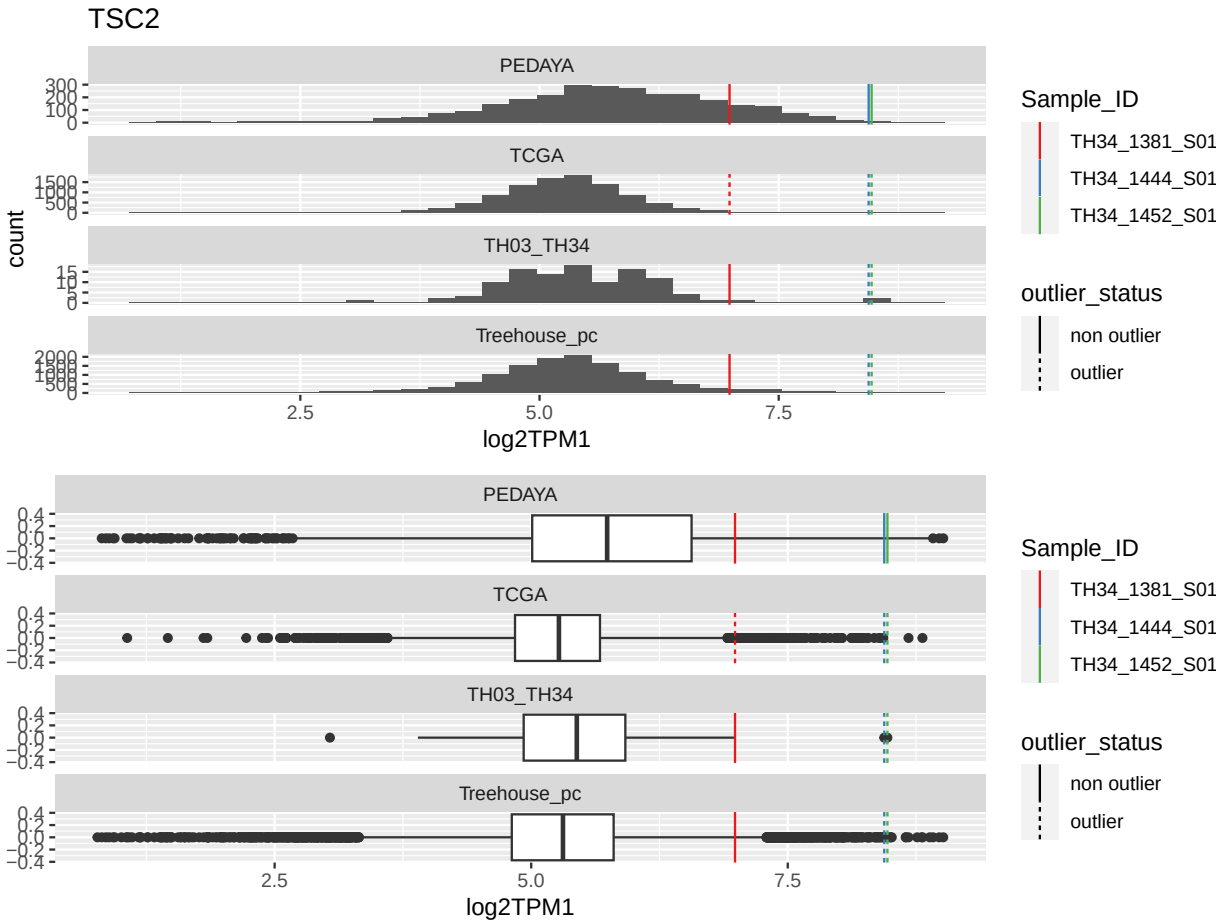
```
##
## [[52]]
```



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

cohort	q25	median	q75	IQR	up_outlier_threshold
PEDAYA	2.93	3.99	5.29	2.36	8.82
TCGA	3.15	3.83	4.46	1.31	6.43
TH03_TH34	2.74	3.55	4.22	1.48	6.44
Treehouse_pc	3.14	3.89	4.63	1.49	6.87

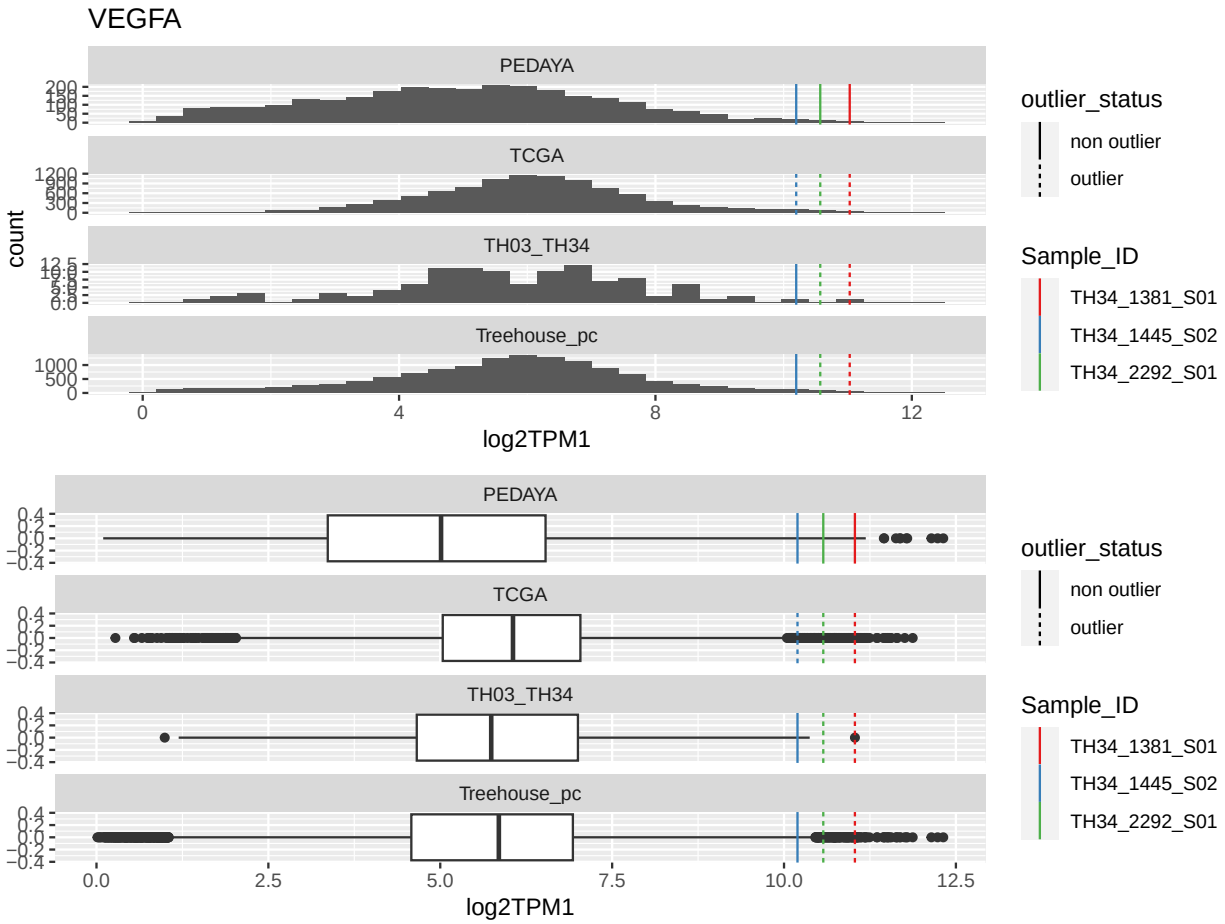
```
##
## [[53]]
```



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

cohort	q25	median	q75	IQR	up_outlier_threshold
PEDAYA	5.01	5.74	6.56	1.55	8.89
TCGA	4.84	5.27	5.67	0.83	6.91
TH03_TH34	4.93	5.44	5.92	0.99	7.40
Treehouse_pc	4.81	5.31	5.80	0.99	7.29

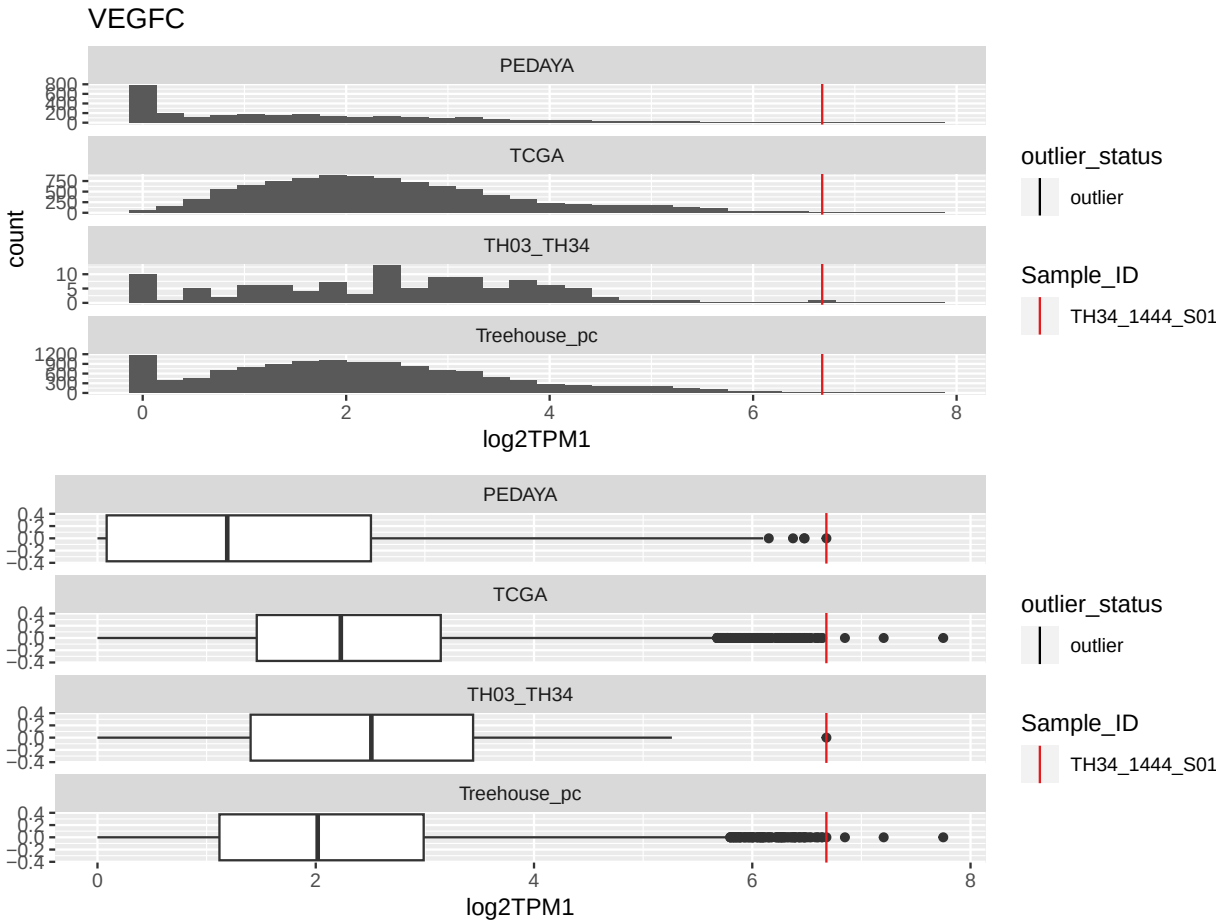
```
##
## [[54]]
```



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_2292_S01	VEGFA	FALSE	TRUE	TRUE	TRUE	10.571

cohort	q25	median	q75	IQR	up_outlier_threshold
PEDAYA	3.36	5.01	6.53	3.17	11.28
TCGA	5.04	6.06	7.04	2.00	10.04
TH03_TH34	4.66	5.74	7.00	2.34	10.52
Treehouse_pc	4.58	5.85	6.93	2.35	10.45

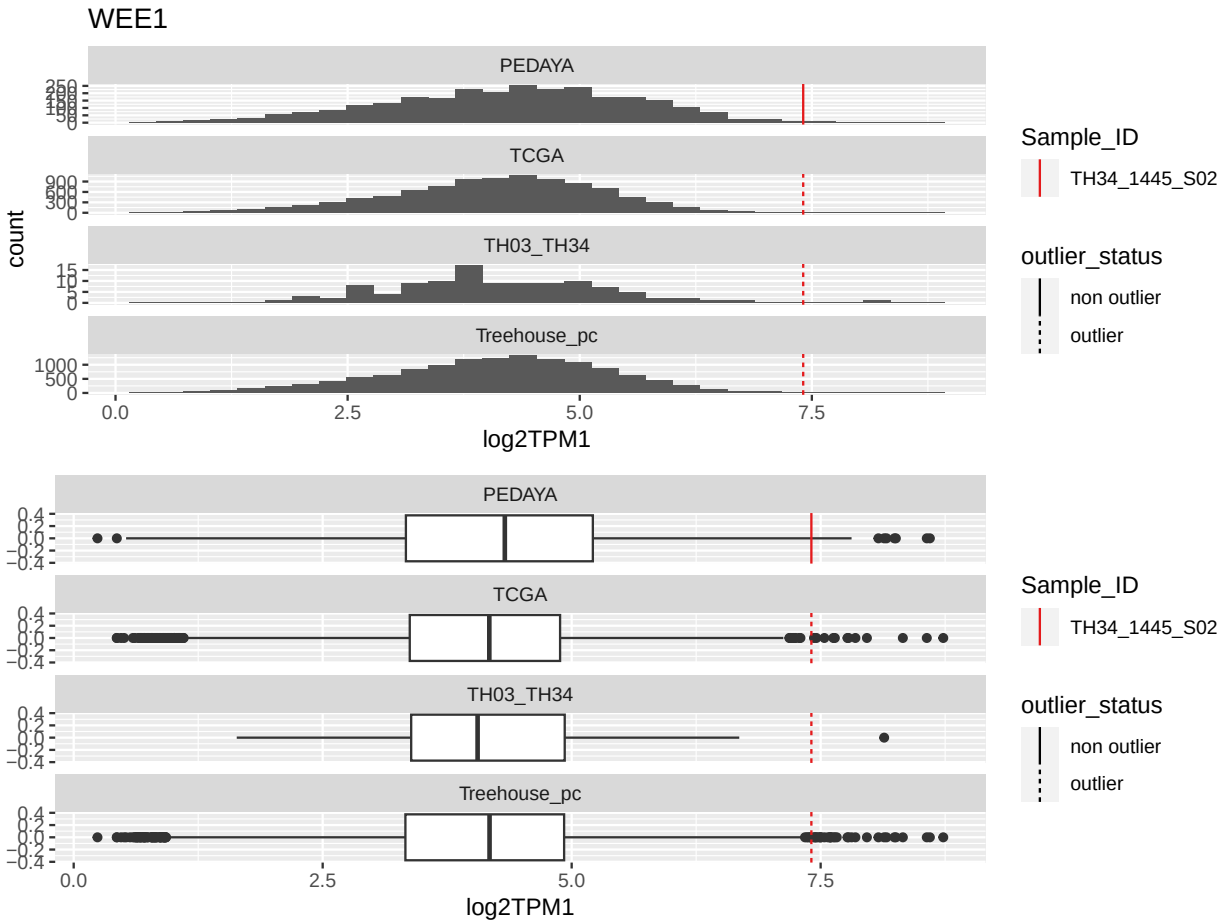
```
##
## [[55]]
```



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

cohort	q25	median	q75	IQR	up_outlier_threshold
PEDAYA	0.08	1.19	2.51	2.42	6.14
TCGA	1.46	2.23	3.15	1.69	5.68
TH03_TH34	1.40	2.51	3.44	2.04	6.50
Treehouse_pc	1.12	2.02	2.99	1.87	5.80

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



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

cohort	q25	median	q75	IQR	up_outlier_threshold
PEDAYA	3.34	4.33	5.21	1.88	8.03
TCGA	3.37	4.17	4.88	1.51	7.15
TH03_TH34	3.39	4.06	4.93	1.54	7.24
Treehouse_pc	3.33	4.18	4.92	1.59	7.31

table for annotating TCGA vs Treehouse pc

```
outlier_table_for_annotation <- outliers %>%
    select(Sample_ID, gene, comparison_cohort) %>%
    mutate(found = TRUE) %>%
    pivot_wider(names_from = comparison_cohort,
                values_from = found,
                values_fill = FALSE) %>%
  arrange(gene) %>%
  mutate(TCGA_not_treehouse_pc = TCGA & ! Treehouse_pc) %>%
  select(gene, Sample_ID, TCGA, Treehouse_pc, TCGA_not_treehouse_pc, everything())

write_tsv(outlier_table_for_annotation, "../gather_input_data/comparison_to_non_CARE_cohorts/TCGA_vs_TH
```