Day 2 – 04 Full Heatmap

Seminar plotting walk-through

This notebook takes `dataset3_subset.csv` / `dataset3_subset_long.csv` (all genomes) and gradually moves from a plain heatmap to a more publication-ready version. Each step adds one feature so you can see the incremental effect.

1. Packages, paths, and helpers

```
library(ComplexHeatmap)
library(circlize)
subset_path <- file.path('..', 'data', 'dataset3_subset.csv')
long_path <- file.path('..', 'data', 'dataset3_subset_long.csv')
pdf_path <- file.path('..', 'pdf', '04_full_heatmap.pdf')
na_color <- '#dcdcdc'
```

2. Load data and basic NA report

```
wide_df <- read.csv(subset_path, check.names = FALSE, stringsAsFactors = FALSE)
long_df <- read.csv(long_path, check.names = FALSE, stringsAsFactors = FALSE)
cat('Wide rows x cols:', nrow(wide_df), ncol(wide_df), '\n')
```

```
## Wide rows x cols: 71 67
```

```
cat('Long rows x cols:', nrow(long_df), ncol(long_df), '\n')
```

```
## Long rows x cols: 4544 7
```

```
na_total <- sum(is.na(long_df$value))
cat('Total NA entries in value:', na_total, '\n')
```

```
## Total NA entries in value: 443
```

```
if (na_total > 0) {
  na_table <- with(long_df, tapply(value, list(mouse_id, day), function(x) sum(is.na(x))))
  print(na_table)
}
```

```
##        0  4  9 14 18 23 30 37 44 49 53 58 63 67 72 79
## 1683   0  1  2  3  4  4  2  0  1  2  1  1  1  0  1  1
## 1688  15  1  1  2  3  0  1  1  0  0  0  1  0  0  1  1
## 1692   0 47 24  0 46 20  6  0  1  9 38  0  0 28  0  0
## 1699   0 23 37  0 24 13  0  0  1  1 23 13  1 23 13  1
```
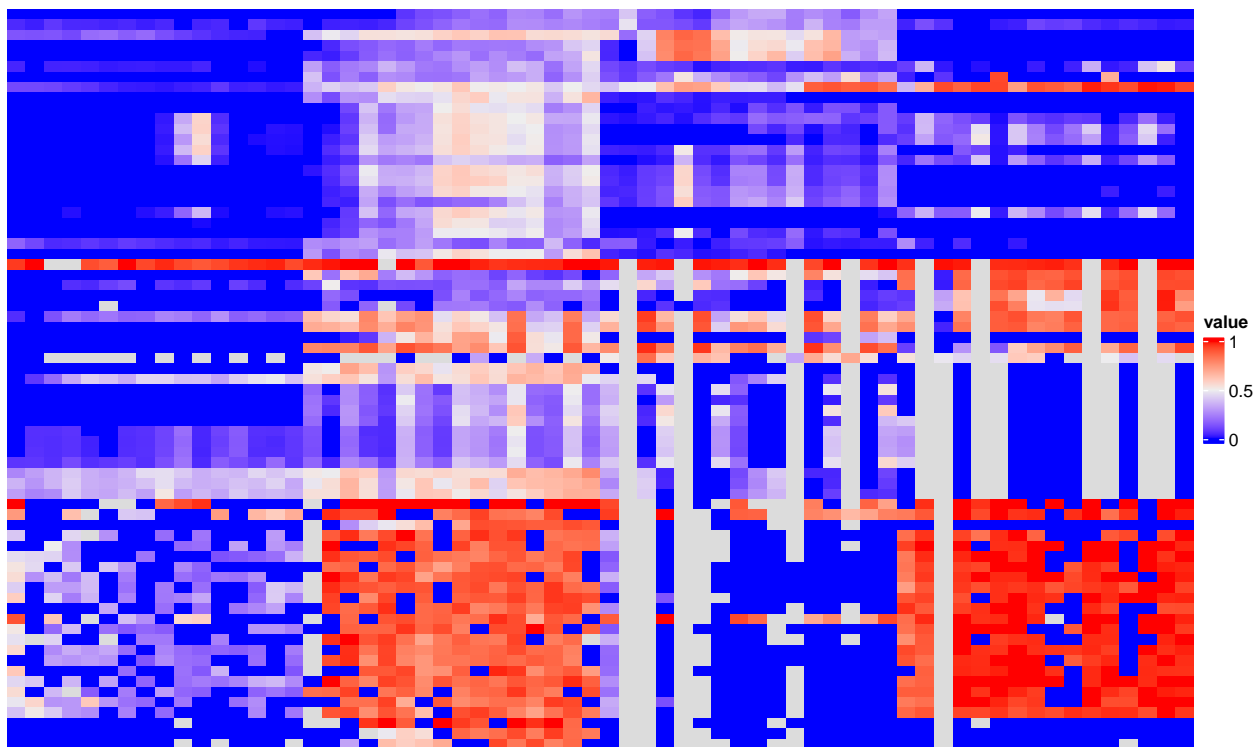
3. Base heatmap (no annotations)

```
sample_cols <- setdiff(names(wide_df), c('Genome', 'snp_id', 'Position'))
mat <- as.matrix(wide_df[, sample_cols])
```

1

```r
mode(mat) <- 'numeric'
rownames(mat) <- paste(wide_df$Genome, wide_df$snp_id, sep = ' | ')
colnames(mat) <- sample_cols

ht_base <- Heatmap(
  mat,
  name = 'value',
  cluster_rows = FALSE,
  cluster_columns = FALSE,
  show_row_names = FALSE,
  show_column_names = FALSE,
  na_col = na_color
)
draw(ht_base)
```
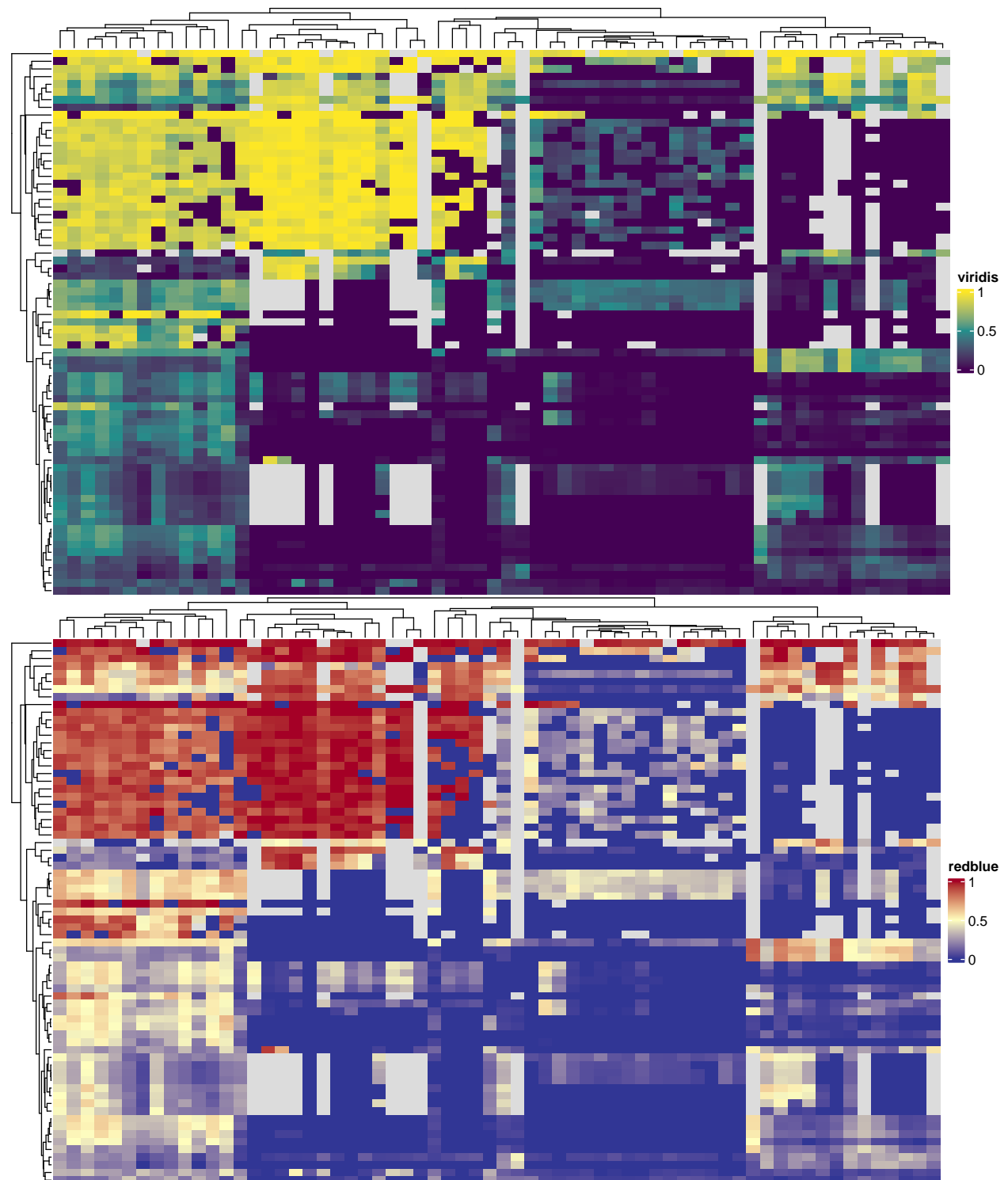

4. Explore color scales

Generate three quick variants to see how different palettes affect the look.

```r
mins <- min(mat, na.rm = TRUE)
maxs <- max(mat, na.rm = TRUE)
mids <- (mins + maxs) / 2
palettes <- list(
  viridis = circlize::colorRamp2(c(mins, mids, maxs), viridisLite::viridis(3)),
  redblue = circlize::colorRamp2(c(mins, mids, maxs), c('#313695', '#ffffbf', '#a50026')),
  grayscale = circlize::colorRamp2(c(mins, mids, maxs), c('#f7f7f7', '#cccccc', '#252525'))
)
for (nm in names(palettes)) {
  ht_tmp <- Heatmap(mat, name = nm, col = palettes[[nm]], show_row_names = FALSE, show_column_names = F
  grid::grid.newpage()
```
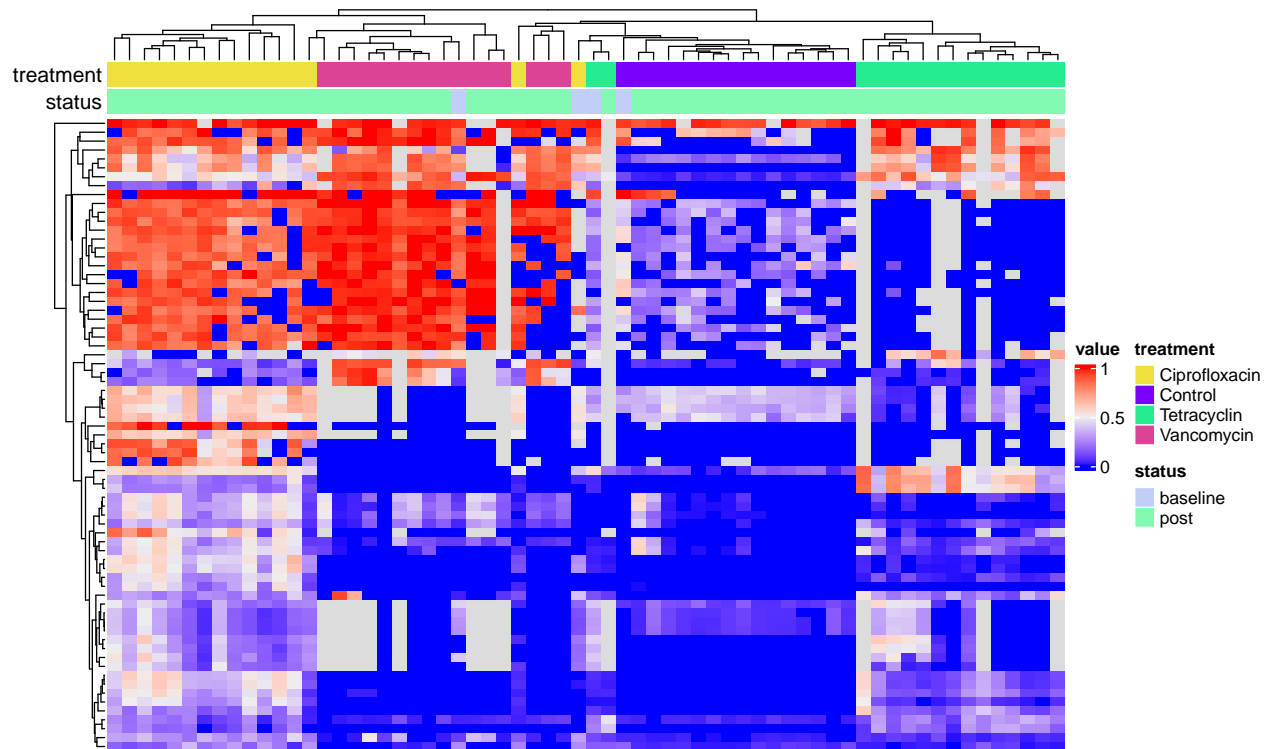
```
draw(ht_tmp)
}
```
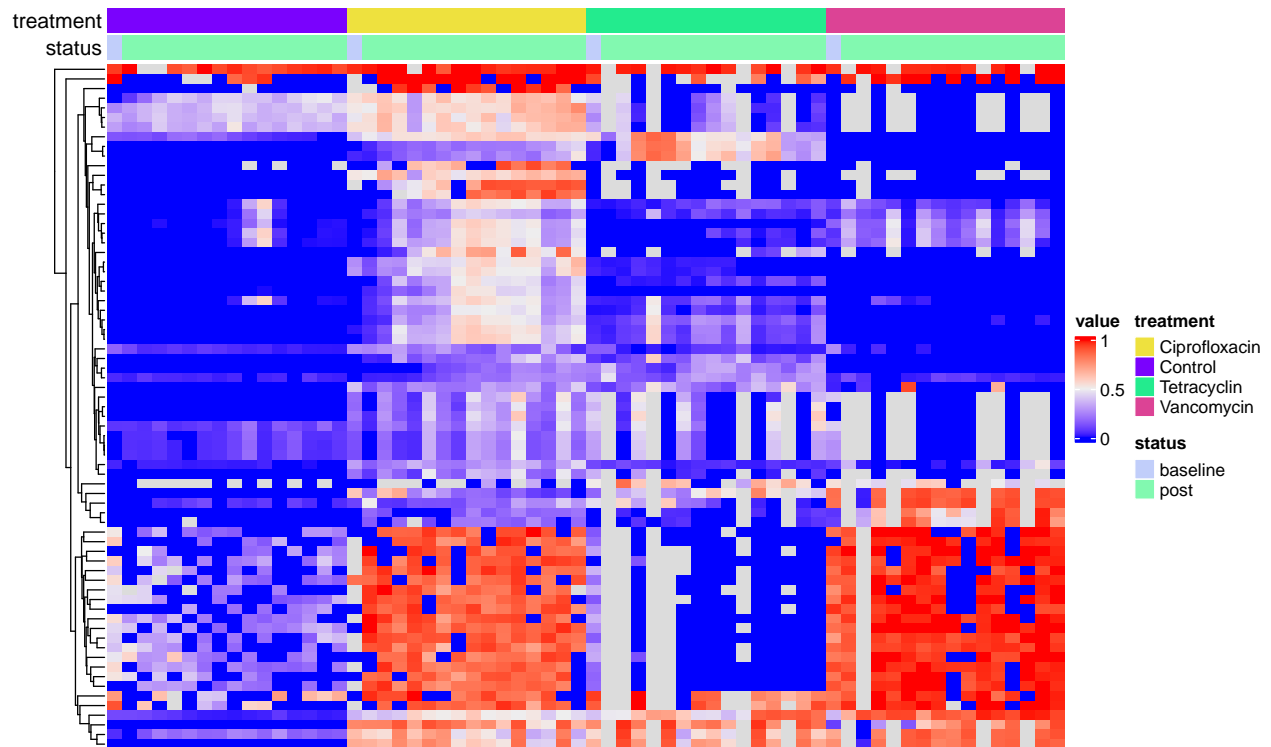
5. Add treatment and baseline/post annotations

```r
sample_meta <- unique(long_df[, c('mouse_id', 'day', 'treatment_group')])
sample_meta$sample_id <- paste(sample_meta$mouse_id, sample_meta$day, sep = '-')
sample_meta <- sample_meta[match(colnames(mat), sample_meta$sample_id), ]
sample_meta$post_ab <- ifelse(sample_meta$day == 0, 'baseline', 'post')
col_ann <- HeatmapAnnotation(
  treatment = sample_meta$treatment_group,
  status = sample_meta$post_ab,
  annotation_name_side = 'left'
)

ht_ann <- Heatmap(
  mat,
  name = 'value',
  show_row_names = FALSE,
  show_column_names = FALSE,
  na_col = na_color,
  top_annotation = col_ann
)
draw(ht_ann)
```
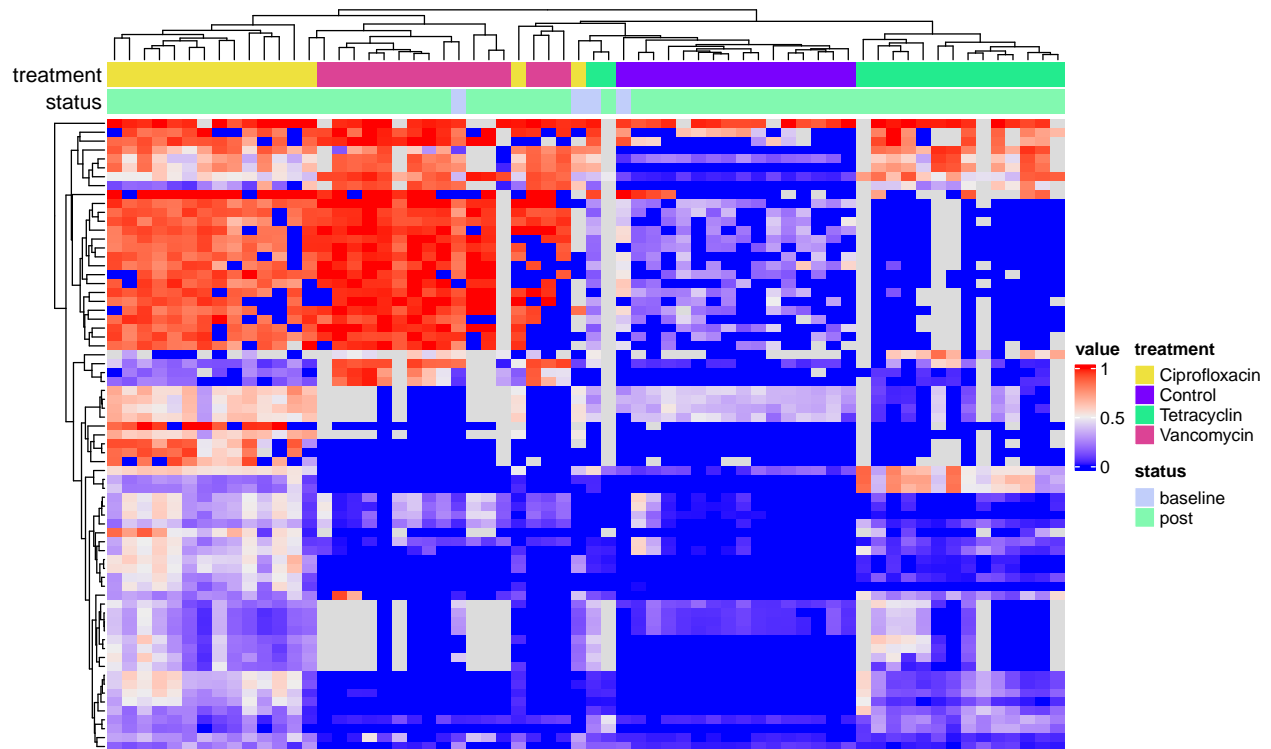
6. Order columns (mouse/day) and cluster rows

```
order_idx <- order(sample_meta$mouse_id, sample_meta$day)
mat_ordered <- mat[, order_idx]
row_dend <- hclust(dist(mat_ordered), method = 'average')
ht_ordered <- Heatmap(
  mat_ordered,
  name = 'value',
  cluster_rows = as.dendrogram(row_dend),
  cluster_columns = FALSE,
  show_row_names = FALSE,
  show_column_names = FALSE,
  na_col = na_color,
  top_annotation = col_ann[order_idx]
)
draw(ht_ordered)
```

7. Row filtering example (top variance)

```r
row_var <- apply(mat, 1, var, na.rm = TRUE)
keep_idx <- order(row_var, decreasing = TRUE)[seq_len(min(100, nrow(mat)))]
mat_var <- mat[keep_idx, ]
ht_var <- Heatmap(
  mat_var,
  name = 'value',
  show_row_names = FALSE,
  show_column_names = FALSE,
  na_col = na_color,
  top_annotation = col_ann
)
draw(ht_var)
```
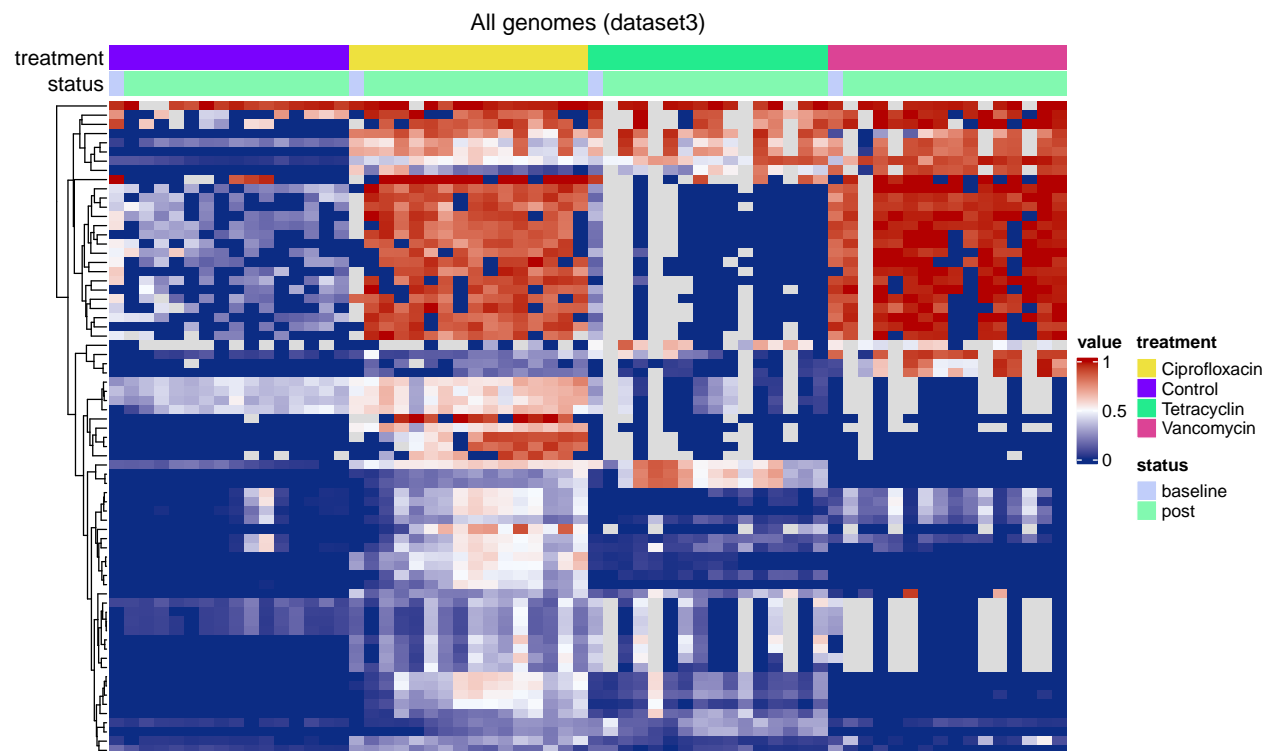
8. Final polish and PDF export

Here we combine ordering, clustering, and annotations into one figure and save it.

```r
ht_final <- Heatmap(
  mat_ordered,
  name = 'value',
  col = circlize::colorRamp2(c(mins, mids, maxs), c('#0c2c84', '#f7fbff', '#b30000')),
  top_annotation = col_ann[order_idx],
  cluster_rows = TRUE,
  cluster_columns = FALSE,
  show_row_names = FALSE,
  show_column_names = FALSE,
  na_col = na_color,
  column_title = 'All genomes (dataset3)'
)

draw(ht_final)
```

All genomes (dataset3)

```
pdf(pdf_path, width = 11, height = 7)
draw(ht_final)
dev.off()
cat('Saved final heatmap to', pdf_path, '\n')
```