Day 2 – 04 Full Heatmap (Exercises Solution)

Seminar reference solution

This solution notebook offers one way to complete the full heatmap exercise. Feel free to tweak palettes, ordering, or filtering thresholds to suit your teaching needs.

1. Packages, paths, helpers

```
suppressPackageStartupMessages({
  library(ComplexHeatmap)
  library(circlize)
  library(viridisLite)
})
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_exercise.pdf')
na_color <- '#dcdcdc'
```

2. Load data and 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('NA count (value):', na_total, '\n')
```

```
## NA count (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)
}
```

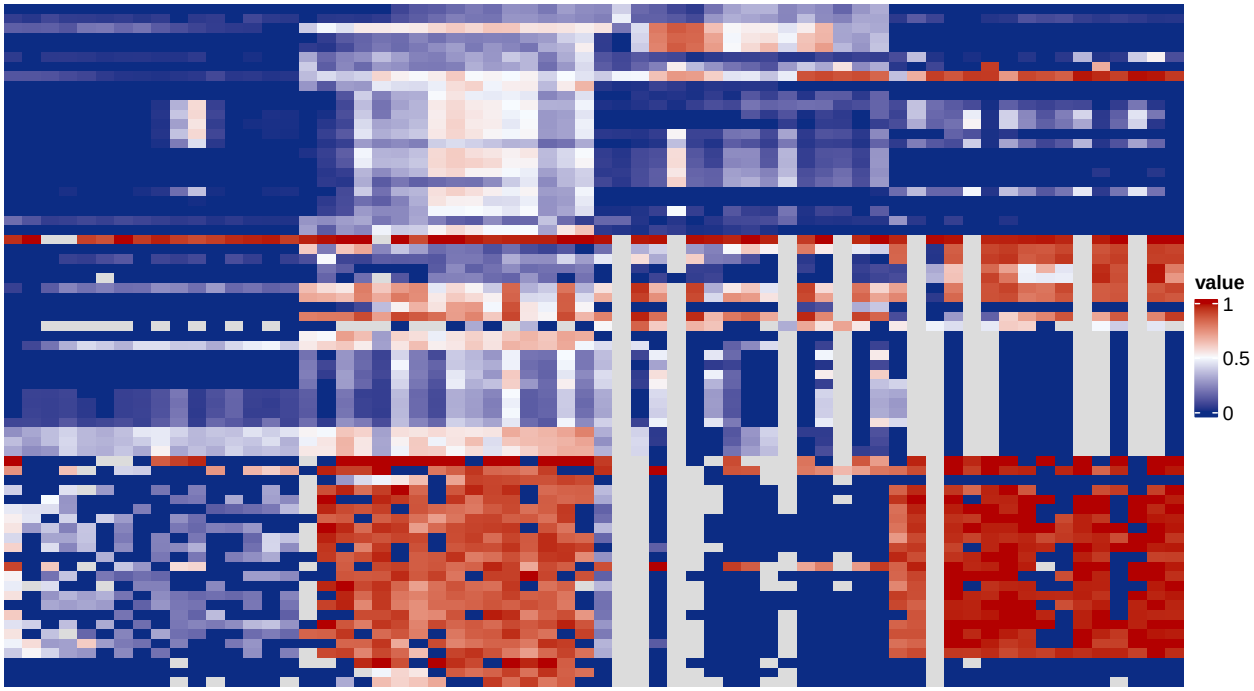
```
##       0  4  9 14 18 23 30 37 44 49 53 58 63 67 72 79
## 1683  0  1  2  3  4  4  2  0  1  2  1  1  1  0  1  1
## 1688 15  1  1  2  3  0  1  1  0  0  0  1  0  0  1  1
## 1692  0 47 24  0 46 20  6  0  1  9 38  0  0 28  0  0
## 1699  0 23 37  0 24 13  0  0  1  1 23 13  1 23 13  1
```

3. Matrix + metadata

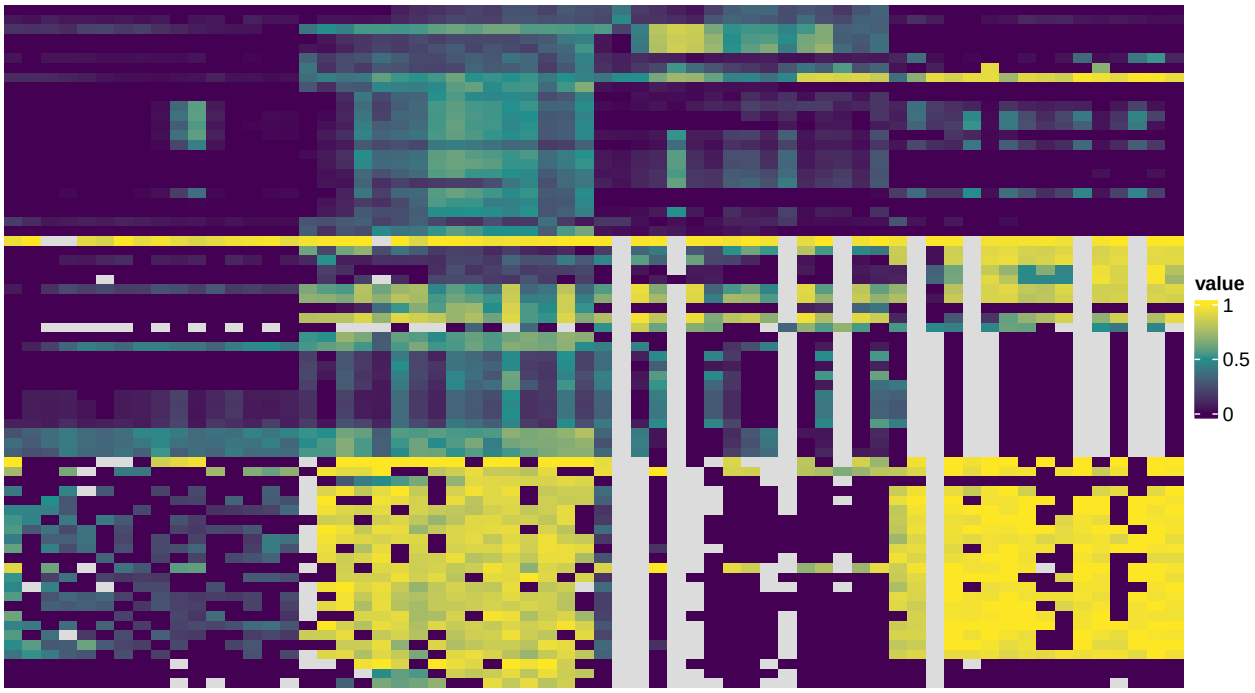
```
sample_cols <- setdiff(names(wide_df), c('Genome','snp_id','Position'))
mat <- as.matrix(wide_df[, sample_cols])
mode(mat) <- 'numeric'
rownames(mat) <- paste(wide_df$Genome, wide_df$snp_id, sep = ' | ')

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), ]
stopifnot(identical(colnames(mat), sample_meta$sample_id))
```

4. Baseline heatmap + palettes

```
mins <- min(mat, na.rm = TRUE)
maxs <- max(mat, na.rm = TRUE)
mids <- (mins + maxs) / 2
palette_blue_red <- circlize::colorRamp2(c(mins, mids, maxs), c('#0c2c84','#f7fbff','#b30000'))
palette_viridis <- circlize::colorRamp2(c(mins, mids, maxs), viridisLite::viridis(3))

ht_blue <- Heatmap(mat, name = 'value', col = palette_blue_red, na_col = na_color,
                   cluster_rows = FALSE, cluster_columns = FALSE,
                   show_row_names = FALSE, show_column_names = FALSE)
ht_viridis <- Heatmap(mat, name = 'value', col = palette_viridis, na_col = na_color,
                      cluster_rows = FALSE, cluster_columns = FALSE,
                      show_row_names = FALSE, show_column_names = FALSE)

ht_blue
```

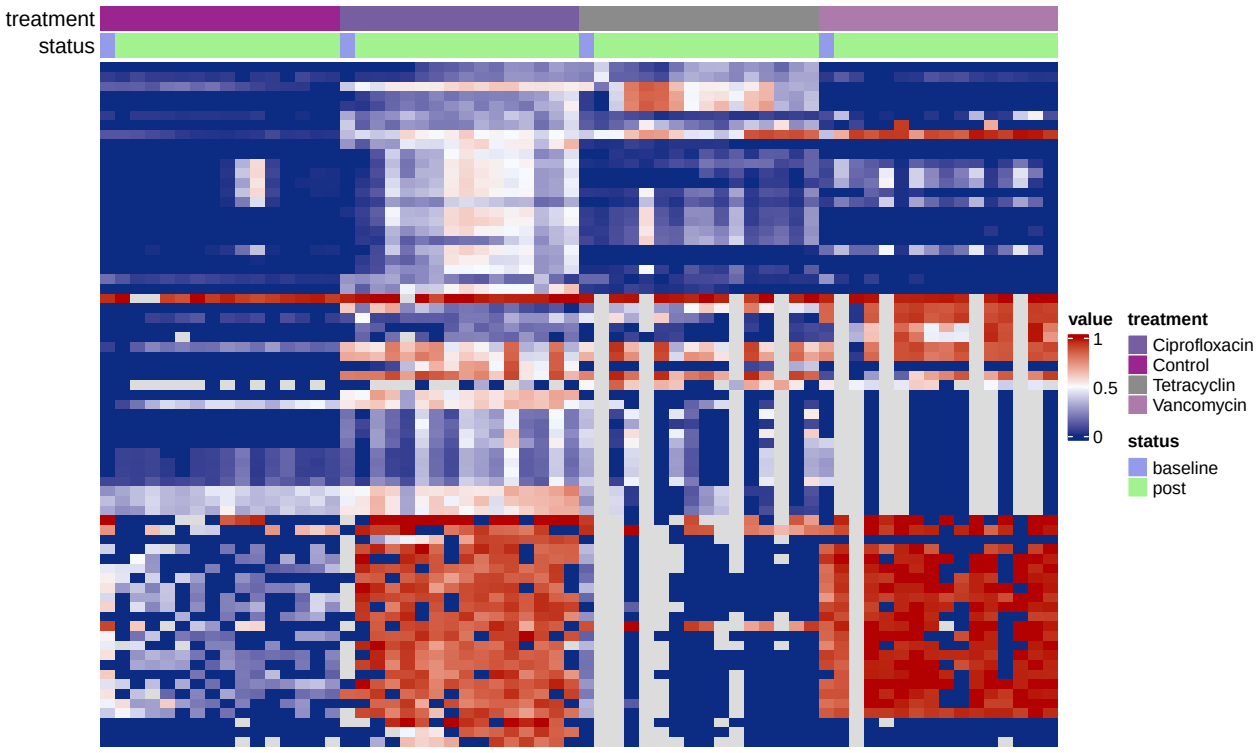


```
ht_viridis
```



5. Annotations + ordering

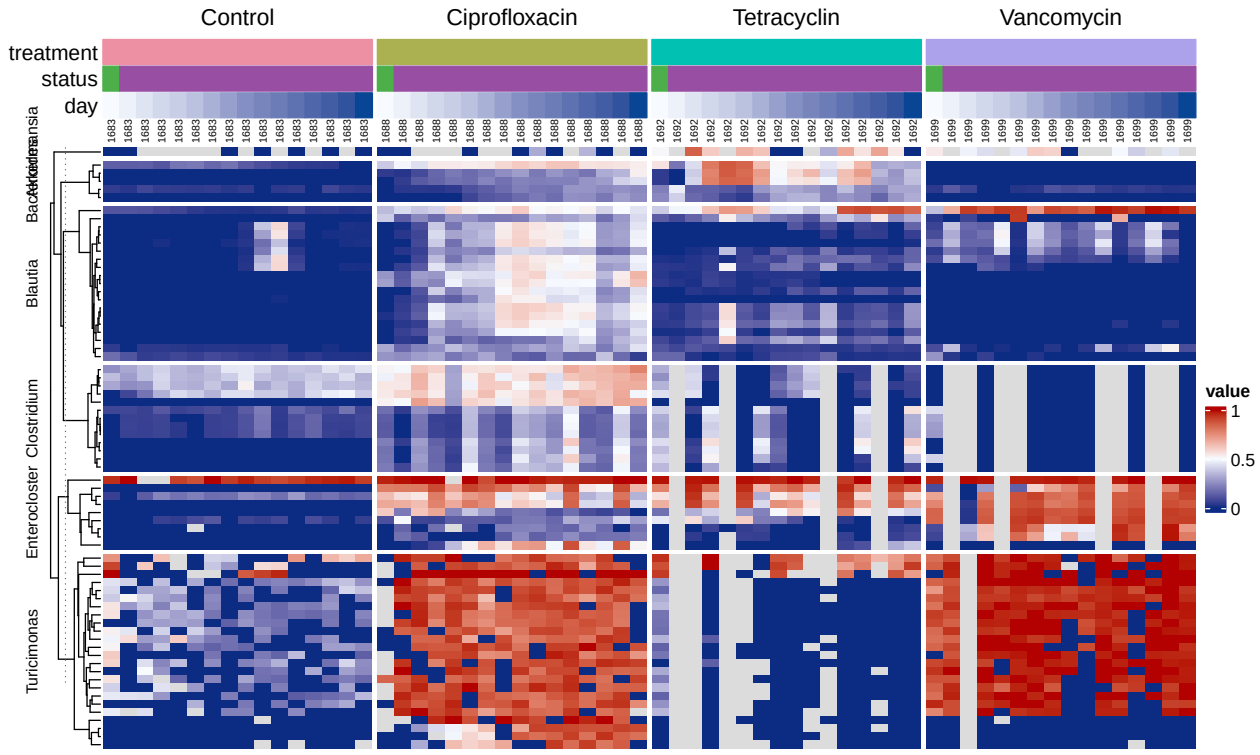
```
sample_meta$post_ab <- ifelse(sample_meta$day == 0, 'baseline', 'post')
col_ann <- HeatmapAnnotation(
  treatment = sample_meta$treatment_group,
  status = sample_meta$post_ab,
  annotation_name_side = 'left'
)
order_idx <- order(sample_meta$mouse_id, sample_meta$day)
mat_ordered <- mat[, order_idx]
col_ann_ordered <- col_ann[order_idx]

Heatmap(
  mat_ordered,
  name = 'value',
  col = palette_blue_red,
  top_annotation = col_ann_ordered,
  cluster_rows = FALSE,
  cluster_columns = FALSE,
  show_row_names = FALSE,
  show_column_names = FALSE,
  na_col = na_color
)
```



6. Annotation enhancements

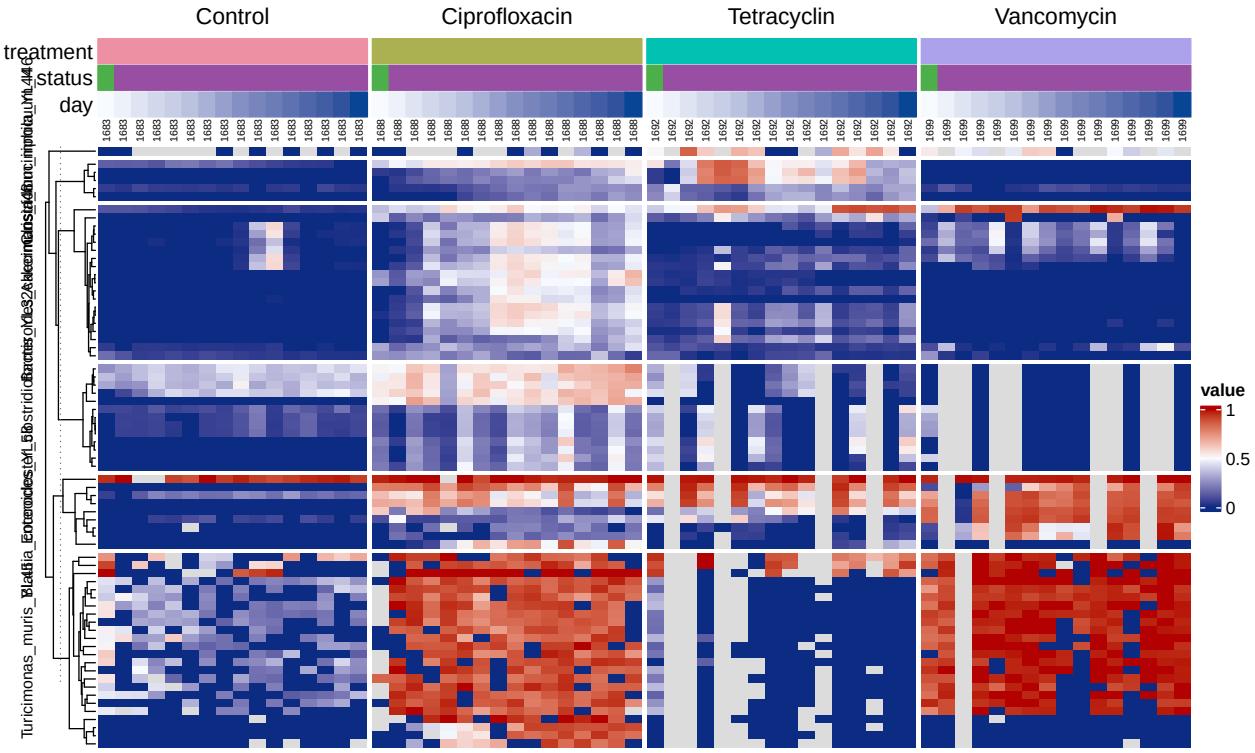
```
treatment_levels <- unique(sample_meta$treatment_group)
treatment_cols <- grDevices::hcl.colors(length(treatment_levels), palette = "Set2")
names(treatment_cols) <- treatment_levels
status_cols <- c(baseline = '#4daf4a', post = '#984ea3')
day_col_fun <- circlize::colorRamp2(range(sample_meta$day), c('#f7fbff','#084594'))
col_ann_rich <- HeatmapAnnotation(
  treatment = anno_simple(sample_meta$treatment_group, col = treatment_cols),
  status = anno_simple(sample_meta$post_ab, col = status_cols),
  day = anno_simple(sample_meta$day, col = day_col_fun),
  mouse = anno_text(sample_meta$mouse_id, rot = 90, gp = grid::gpar(fontsize = 6)),
  annotation_name_side = 'left'
)
col_ann_rich_ordered <- col_ann_rich[order_idx]
column_split <- factor(sample_meta$treatment_group[order_idx], levels = treatment_levels)
row_split <- factor(wide_df$Genome, levels = unique(wide_df$Genome))
row_titles <- sub('_.*', '', levels(row_split))

Heatmap(
  mat_ordered,
  name = 'value',
  col = palette_blue_red,
  top_annotation = col_ann_rich_ordered,
  cluster_rows = TRUE,
  cluster_columns = FALSE,
  column_split = column_split,
  show_row_names = FALSE,
  show_column_names = FALSE,
```

```
  na_col = na_color,
  row_split = row_split,
  row_title = row_titles,
  row_title_gp = grid::gpar(fontsize = 9)
)
```



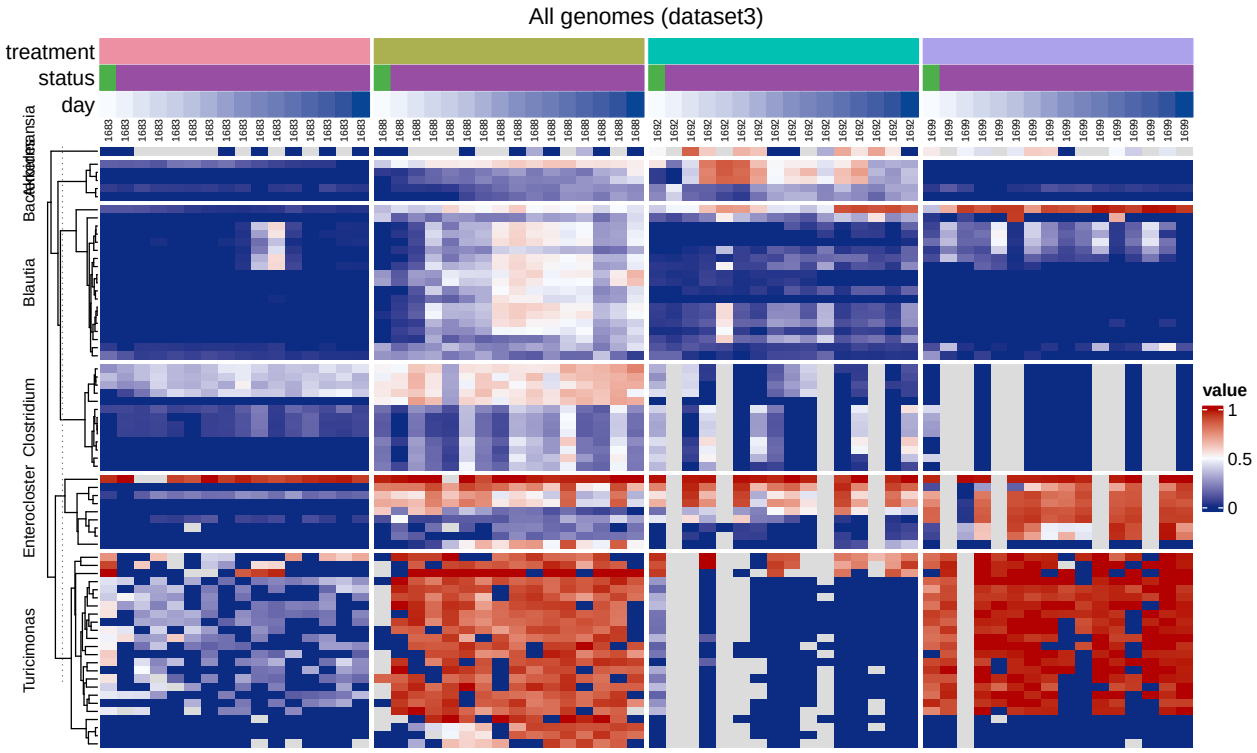
7. Row variance filter

```
row_var <- apply(mat, 1, var, na.rm = TRUE)
keep_idx <- order(row_var, decreasing = TRUE)[seq_len(min(100, nrow(mat)))]
mat_topvar <- mat[keep_idx, order_idx]
row_split_top <- droplevels(row_split[keep_idx])
Heatmap(
  mat_topvar,
  name = 'value',
  col = palette_blue_red,
  top_annotation = col_ann_rich_ordered,
  cluster_rows = TRUE,
  cluster_columns = FALSE,
  column_split = column_split,
  show_row_names = FALSE,
  show_column_names = FALSE,
  na_col = na_color,
  row_split = row_split_top,
  row_title_gp = grid::gpar(fontsize = 9)
)
```



8. Final heatmap + PDF export

```
ht_final <- Heatmap(
  mat_ordered,
  name = 'value',
  col = palette_blue_red,
  top_annotation = col_ann_rich_ordered,
  cluster_rows = TRUE,
  cluster_columns = FALSE,
  column_split = column_split,
  show_row_names = FALSE,
  show_column_names = FALSE,
  na_col = na_color,
  row_split = row_split,
  row_title = row_titles,
  row_title_gp = grid::gpar(fontsize = 9),
  column_title = 'All genomes (dataset3)'
)
draw(ht_final)
```




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

9. Notes

- Palette choice: the blue-white-red ramp emphasizes deviations from mid values.
- Column ordering by mouse/day reveals antibiotic pulses more clearly than the CSV default.
- Variance filtering is helpful when presenting in class; it trims the figure to the most dynamic SNPs and speeds up PDF export.
- Extra annotations (day gradient + mouse labels + row splits) plus column splits by treatment mirror the annotation tutorial and keep the story tied to the experimental design.