Figure 5

This notebook reproduces the main analyses from figure 5 of the manuscript. First we import the necessary packages.

```
library(voxhunt)
library(tidyverse)
library(Seurat)
library(ChIPseeker)
library(GenomicRanges)
library(TxDb.Hsapiens.UCSC.hg19.knownGene)
library(org.Hs.eg.db)
library(Matrix.utils)
```

Now we load the data. The loaded seurat object contains C1 scATAC-seq data from organoid cortex.

```
load_aba_data('voxhunt_data/')
ctx_atac <- read_rds('c1_ctx_srt.rds')
print(ctx_atac)
```

```
## An object of class Seurat
## 73639 features across 506 samples within 1 assay
## Active assay: peaks (73639 features, 0 variable features)
```

Now we need to annotate the peaks using ChIPseeker. However, there should now be more recent methods available

```
ctx_acc <- GetAssayData(ctx_atac, slot='counts')
peak_sums <- Matrix::rowSums(ctx_acc)
det_peaks <- names(peak_sums)[peak_sums>0]

peak_meta <- tibble(peak=det_peaks)
peak_ranges <- peak_meta %>%
    mutate(
        chrom=as.character(
            str_replace(peak, '(chr[0-9XY]+)-\\d+-\\d+', '\\1')),
        start=as.numeric(
            str_replace(peak, 'chr[0-9XY]+-(\\d+)-\\d+', '\\1')),
        end=as.numeric(
            str_replace(peak, 'chr[0-9XY]+-\\d+-(\\d+)', '\\1'))
        ) %>%
    dplyr::select(peak, chrom, start, end) %>%
    GRanges()

peak_annot <- annotatePeak(
    peak_ranges,
```

```
    TxDb = TxDb.Hsapiens.UCSC.hg19.knownGene,
    annoDb = 'org.Hs.eg.db',
    level = 'gene'
)
```

Now we can summarize the peaks to TSS and Promotor regions. We note that there are now more modern ways to do this, e.g. using the `Signac` package.

```
peak_annot_df <- peak_annot@anno %>%
    as_tibble() %>%
    dplyr::select(peak, 'gene'=SYMBOL, annotation) %>%
    mutate(at_gene = str_detect(annotation, 'Intron|Exon|Promoter')) %>%
    filter(!is.na(gene))
tss_peaks <- peak_annot_df %>%
    filter(str_detect(annotation, 'Promoter'))
peak_sums_tss <- aggregate.Matrix(
    as.matrix(ctx_acc[tss_peaks$peak, ]), groupings = tss_peaks$gene)
```

Now we select some markers and map the data

```
marker_df <- structure_markers('E13', 'custom_3')
marker_genes <- marker_df %>%
    filter(gene%in%rownames(peak_sums_tss)) %>%
    group_by(group) %>%
    top_n(10, auc) %>%
    pull(gene) %>% unique()

enrich_map <- voxel_map(t(peak_sums_tss), 'E13',
    genes_use = marker_genes, groups = ctx_atac$cell_type)
plot_map(enrich_map) & no_legend()
```