Figure 3

This notebook reproduces the main analyses from figure 3 of the manuscript. First we import the necessary packages.

```
library(voxhunt)
library(tidyverse)
library(Seurat)
```

Now we load the data. The loaded seurat object contains the neuronal popultions of the datasets shown in the manuscript. We further subset the ones shown in figure 2.

```
load_aba_data('voxhunt_data/')
neurons <- read_rds('combined_neurons_srt.rds')
neurons <- subset(neurons, orig.ident%in%c('cerebral', 'hCS', 'hSS', 'tanaka_thalamus'))
neurons <- subset(neurons,
    cluster%in%c('mesen_ex_cerebral', 'mesen_in_cerebral', 'ctx_ex_cerebral',
        'ge_in_cerebral', 'dien_ex_cerebral', 'ge_hss', 'ctx_hcs', 'dien_tho')
)
print(unique(neurons$cluster))
```

```
## [1] "mesen_ex_cerebral" "ctx_ex_cerebral"    "dien_ex_cerebral"
## [4] "ge_in_cerebral"     "mesen_in_cerebral" "ctx_hcs"
## [7] "ge_hss"             "dien_tho"
```

We can see that `cluster` already captures the different neuronal types we are interested in. Now we select some structure markers.

```
struct_markers <- structure_markers('E13', 'custom_2')
genes_use <- struct_markers %>%
    group_by(group) %>%
    top_n(15, auc) %>%
    pull(gene) %>% unique()
print(head(genes_use))
```

```
## [1] "TCF7L2"  "NR2F1"   "OTX2"    "NEUROG2" "TENM3"   "CBLN2"
```

Now we run VoxHunt with the selected genes and plot the similarities to brain structures.

```
neuron_voxmap <- voxel_map(
    neurons,
    group_name='cluster',
    genes_use=genes_use,
    pseudobulk_groups=T
)
```

```r
strct_cor <- summarize_groups(neuron_voxmap) %>%
    dplyr::filter(custom_2!='medullary hindbrain') %>%
    dplyr::mutate(struct_name=case_when(
        str_detect(custom_2, 'hindbrain') ~ 'hindbrain',
        str_detect(custom_4, 'septum|subpall|striatum|amygda|telencephalic') ~ 'subpallium',
        str_detect(custom_2, 'telen') ~ 'pallium',
        TRUE ~ custom_2
    )) %>%
    dplyr::mutate(struct_name=factor(struct_name, levels=struct_names)) %>%
    dplyr::arrange(struct_name) %>%
    dplyr::mutate(custom_4=factor(custom_4, levels=unique(.$custom_4))) %>%
    dplyr::group_by(group, custom_4, struct_name) %>%
    dplyr::summarise(corr=mean(corr)) %>%
    dplyr::filter(!str_detect(custom_4, 'telencephalo'))

ggplot(strct_cor, aes(custom_4, corr, fill=struct_name)) +
    geom_bar(stat='identity') +
    facet_wrap(group~., scales = 'free', ncol=2) +
    scale_fill_manual(values=struct_colors) +
    theme_bw() +
    theme(axis.text.x=element_blank(), axis.ticks.x=element_blank())
```
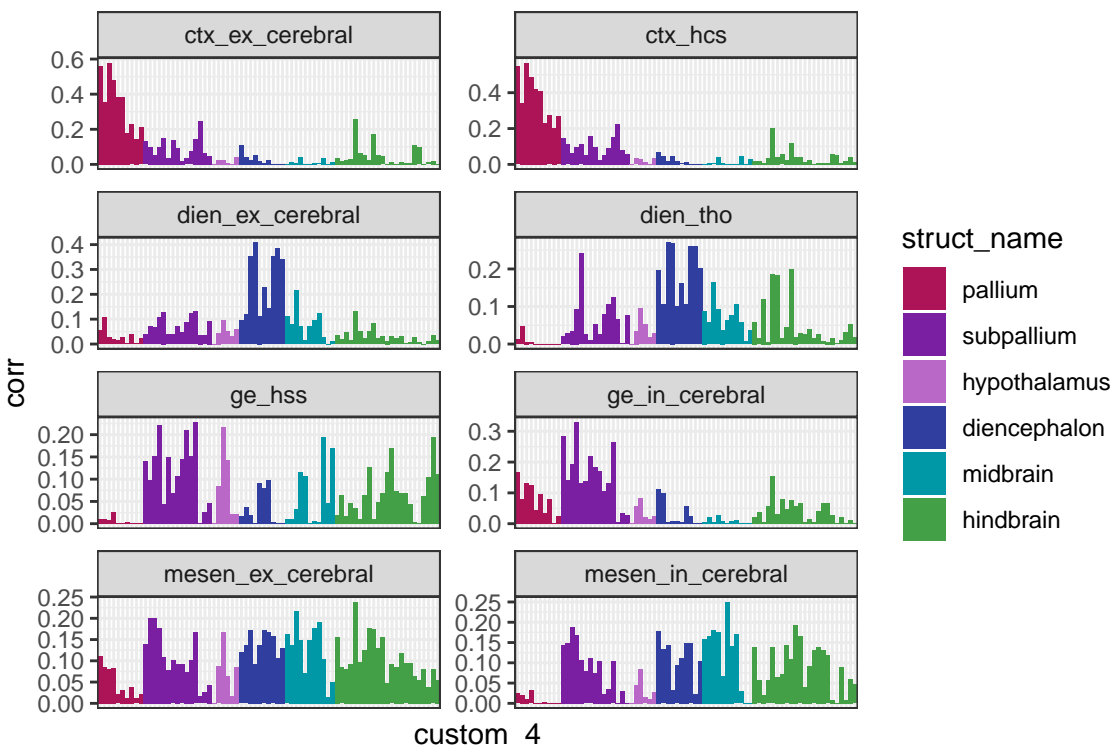

If we further run VoxHunt on a single cell level, we can assign each cell to the maximum correlating structure using `assign_cells()` we can then color the UMAP projection based on this assignment.