

Tph2 Fiber Analysis

Introduction

This notebook takes in exported CellProfiler data along with a metadata table with information about treatment for each mouse and outputs graphs and statistics about the fibers.

Setup

```
library(ggplot2)
```

Warning: package 'ggplot2' was built under R version 4.1.2

```
library(here)
```

here() starts at /Users/rsenft/Documents/GitHub/RebeccaSenft_Projects/002_Dymecki_neuro/Batch_3_fiber

```
library(dplyr)
```

Warning: package 'dplyr' was built under R version 4.1.2

Attaching package: 'dplyr'

The following objects are masked from 'package:stats':

filter, lag

The following objects are masked from 'package:base':

intersect, setdiff, setequal, union

```
library(RColorBrewer)
```

Warning: package 'RColorBrewer' was built under R version 4.1.2

```
library(readr)
```

Warning: package 'readr' was built under R version 4.1.2

```
library(DT)
```

Warning: package 'DT' was built under R version 4.1.2

```
library(knitr)
```

Warning: package 'knitr' was built under R version 4.1.2

```
opts_chunk$set(tidy.opts=list(width.cutoff=I(60)),tidy=TRUE)
```

Get data

Note that the CSV had to be corrected because there is one filename with ROB instead of ROb and because not all the metadata is extracted correctly (inconsistent filenames), so I manually corrected this in Excel using Flash Fill and correcting the misspelling and saved this as Tph2_Per_Image_metadata_corrected.csv.

```
path_to_csv_folder = here("csvs")
save_path <- path_to_csv_folder
img_filename = "Tph2_Per_Image_metadata_corrected.csv"
sc_filename = "Tph2_Per_BorderedCells.csv"

metadata_filename = "Litter-groups-A-through-ii.csv" #note I gave this
↳ headings

# load data and metadata for images
image_df <- read.csv(here(path_to_csv_folder, img_filename),
  check.names = FALSE)
metadata <- read.csv(here(path_to_csv_folder,
  metadata_filename), check.names = FALSE)
```

```

image_df_combined <- merge(image_df, metadata,
  by.x = "Image_Metadata_Litter", by.y = "Litter")

# load data for single cells
sc_df <- read.csv(here(path_to_csv_folder, sc_filename),
  check.names = FALSE)
sc_df_combined <- merge(sc_df, image_df_combined,
  by = "ImageNumber")

```

We also need to compute the average length of a branch, taken as the total skeleton length / (number of trunks + number of non-trunk branches).

```

sc_df_combined <- sc_df_combined %>%
  mutate(BorderedCells_ObjectSkeleton_avg_branch_length =
    ↪ BorderedCells_ObjectSkeleton_TotalObjectSklnLngh_MrphlgclSkln/(BorderedCells_ObjectSklnLngh_MrphlgclSkln
    ↪ +
    BorderedCells_ObjectSkeleton_NumberTrunks_MorphologicalSkeleton))

```

Summarize skeleton features

Summarize for each region, mouse, treatment (region_df) and for each mouse and treatment group (total_df).

```

# Select only columns with fiber measurements and create
# summary table:
cols2summarize <- names(sc_df_combined)[grepl("*ObjectSkeleton*",
  names(sc_df_combined))]
# make Inf NaN
sc_df_combined[sc_df_combined == Inf] <- NaN
# get morphological skeleton features

region_df <- sc_df_combined %>%
  group_by(ExpGroup, Image_Metadata_Mouse,
    Image_Metadata_Region) %>%
  summarize(across(all_of(cols2summarize),
    mean, na.rm = TRUE))

```

`summarise()` has grouped output by 'ExpGroup', 'Image_Metadata_Mouse'. You can override using the `.groups` argument.

```

total_df <- region_df %>%
  group_by(ExpGroup, Image_Metadata_Mouse) %>%
  summarize(across(all_of(cols2summarize), mean,

```

```
na.rm = TRUE))
```

`summarise()` has grouped output by 'ExpGroup'. You can override using the `groups` argument.

Plot skeleton features

Plot the five features by region and by experimental group ExpGroup. First, we set up some plotting colors and descriptions for each variable:

```
pal = c("#FFE66D", "#C2CAE8", "#FF4365", "#00D9C0",
        "#F26419", "#86BBD8")
pal_dark = c("#A38800", "#273568", "#FF6B6B", "#FF4365",
             "#00D9C0", "#F26419", "#86BBD8")

measurements =
  ↪ c("BorderedCells_ObjectSkeleton_NumberBranchEnds_MorphologicalSklttn",
      "BorderedCells_ObjectSkeleton_NumberNonTrunkBranchs_MrphlgclSklttn",
      "BorderedCells_ObjectSkeleton_NumberTrunks_MorphologicalSkeleton",
      "BorderedCells_ObjectSkeleton_TotalObjectSklttnLngth_MrphlgclSklttn",
      "BorderedCells_ObjectSkeleton_avg_branch_length")
desc = c("Number of Branch Termini",
         "Number of Trunk Branches", "Number of Trunks",
         "Total Skeleton Length", "Mean Process Length")
```

Here is the function for our box and whisker plot:

```
plot_boxwhisker <- function(data, measurements,
                             desc, index, x_var) {
  meas <- measurements[index]
  data <- data %>%
    filter(is.finite(get(meas)))
  p <- ggplot(data, aes(x = get(x_var),
                       y = get(meas), fill = ExpGroup)) +
    scale_fill_manual(values = pal) +
    geom_boxplot(position = position_dodge(0.9)) +
    geom_jitter(position = position_dodge(0.9),
               alpha = 0.8, aes(color = ExpGroup)) +
    scale_color_manual(values = pal_dark) +
    stat_summary(fun = mean, geom = "point",
               shape = 18, position = position_dodge(0.9),
               size = 3, color = "firebrick") +
    # geom_jitter(position = position_dodge(0.9),
    # alpha=0.4)+
  labs(title = paste0("Average ", desc[index]),
```

```

    subtitle = "Each dot represents 1 animal",
    caption = "Source: Batch 3",
    x = "Region", y = desc[index]) +
    theme(text = element_text(size = 20))
  return(p)
}

```

Now we apply that function to each of the 5 variables we want to plot:

```

p_list = list()
for (i in 1:length(measurements)) {
  p_list[[i]] <- plot_boxwhisker(region_df,
    measurements, desc, i,
    x_var = "Image_Metadata_Region")
  plot(p_list[[i]])
}

```

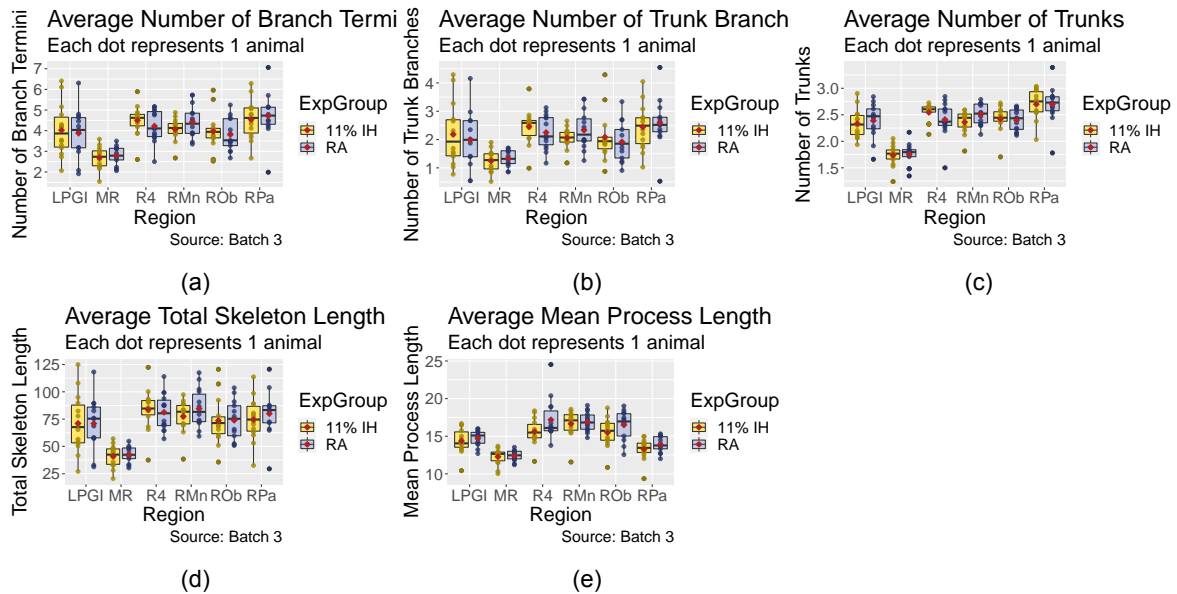


Figure 1: No effect of experimental group on fiber measures.

In examining Mean Process Length, it almost looks like there could be a difference associated with experimental group in R4. We can look at this further by conducting a t-test. RA and IH are not statistically different:

```

# t test
y_var = "BorderedCells_ObjectSkeleton_avg_branch_length"
grp1 <- region_df %>%
  filter(Image_Metadata_Region == "R4", ExpGroup ==

```

```

    "11% IH") %>%
  pull(y_var)
grp2 <- region_df %>%
  filter(Image_Metadata_Region == "R4", ExpGroup ==
    "RA") %>%
  pull(y_var)
t.test(grp1, grp2)

```

Welch Two Sample t-test

```

data:  grp1 and grp2
t = -1.8325, df = 22.148, p-value = 0.08035
alternative hypothesis: true difference in means is not equal to 0
95 percent confidence interval:
 -3.361794  0.207057
sample estimates:
mean of x mean of y
 15.60552  17.18289

```

Now let's just look overall at RA vs. IH mice:

```

p_list = list()
for (i in 1:length(measurements)) {
  p_list[[i]] <- plot_boxwhisker(total_df, measurements,
    desc, i, x_var = "ExpGroup")
  plot(p_list[[i]])
}

# cowplot::plot_grid(plotlist=p_list, ncol=2, scale=1)

```

Instead of viewing each animal as a datapoint, we can also look at the distribution of all cells. Here's our plotting function:

```

plot_distribution <- function(data,
  measurements, desc, index, x_var,
  facetby) {
  p <- ggplot(sc_df_combined,
    aes(x = get(measurements[index]),
      fill = ExpGroup, color = ExpGroup)) +
    geom_density(alpha = 0.7,
      adjust = 1, color = NaN) +
    scale_fill_manual(values = pal) +
    geom_density(alpha = 0.2,
      adjust = 1, fill = NaN) +
    scale_color_manual(values = pal_dark) +

```

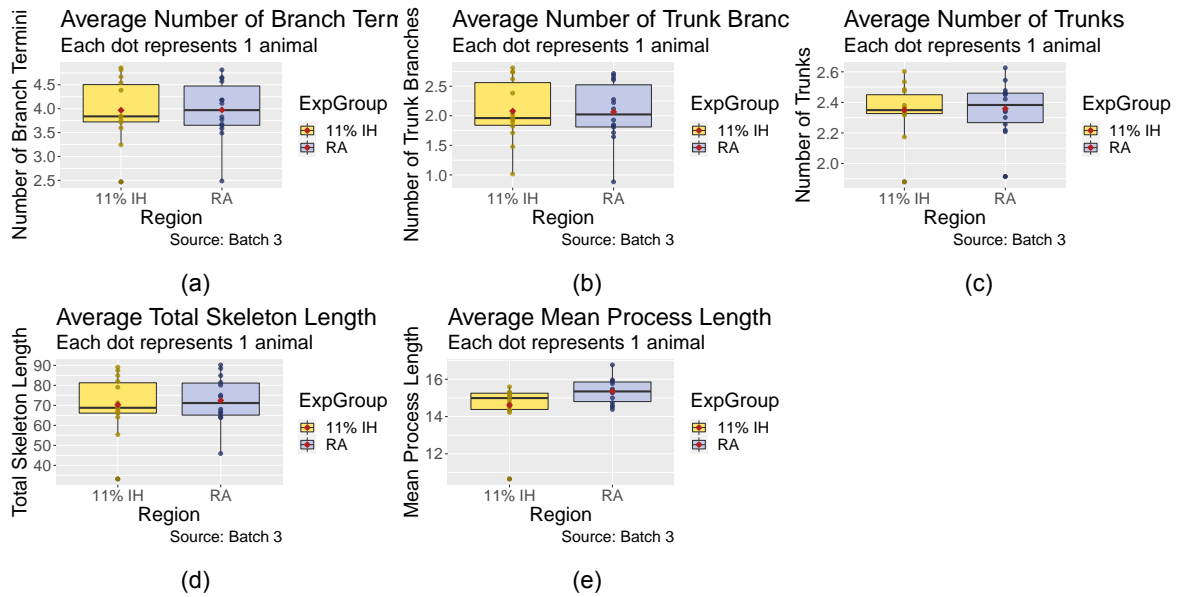


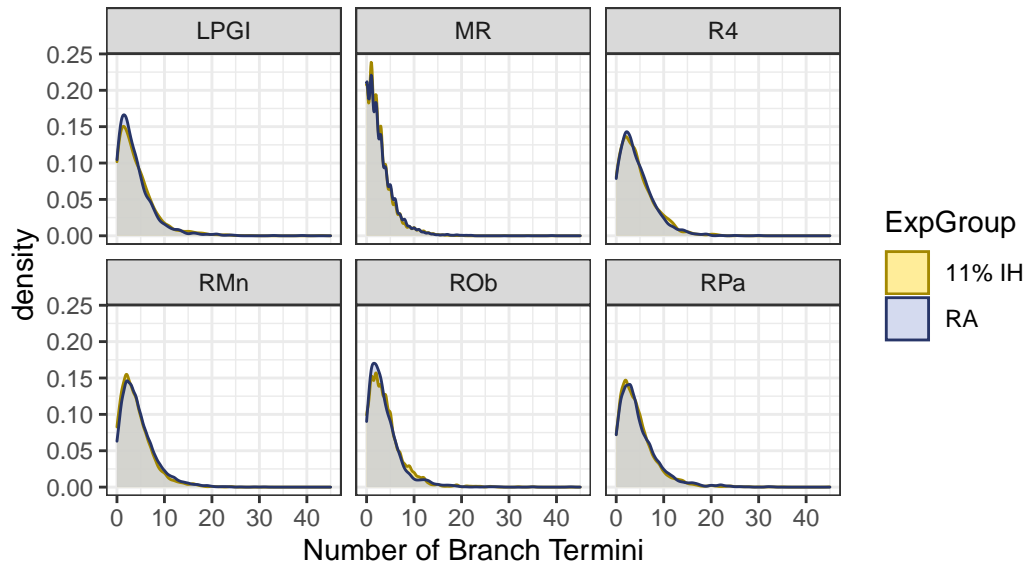
Figure 2: No effect of experimental group on fiber measures.

```
facet_wrap(~get(facetby)) +
labs(title = paste0("Distribution of ",
  desc[index], " by Region, treatment"),
caption = "Source: Batch 3",
x = desc[index]) + theme_bw()
print(p)
}
```

Let's make some distribution plots:

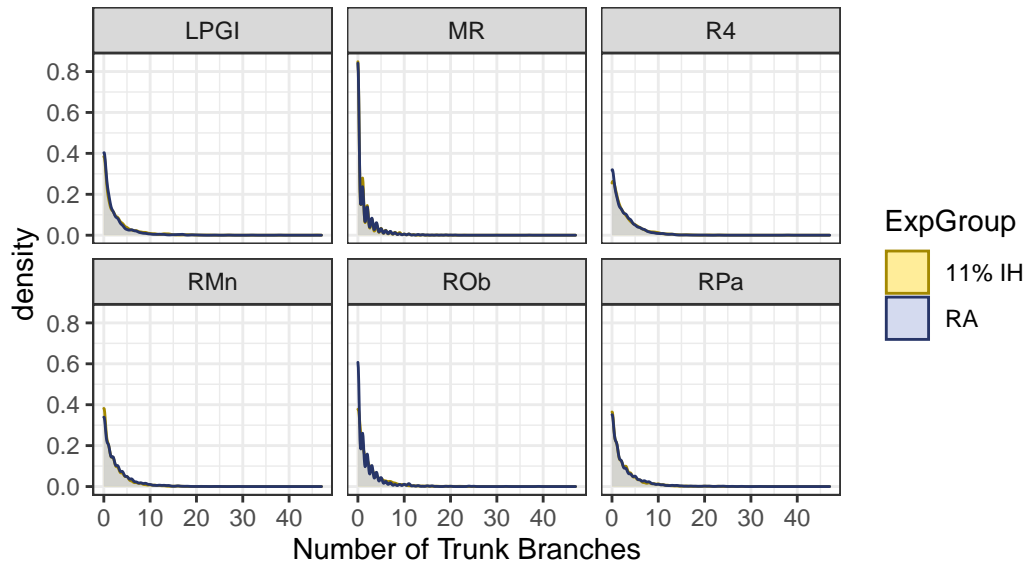
```
for (i in 1:length(measurements)) {
  plot_distribution(sc_df_combined,
    measurements, desc, i, x_var = "ExpGroup",
    facetby = "Image_Metadata_Region")
}
```

Distribution of Number of Branch Termini by Region, treatment



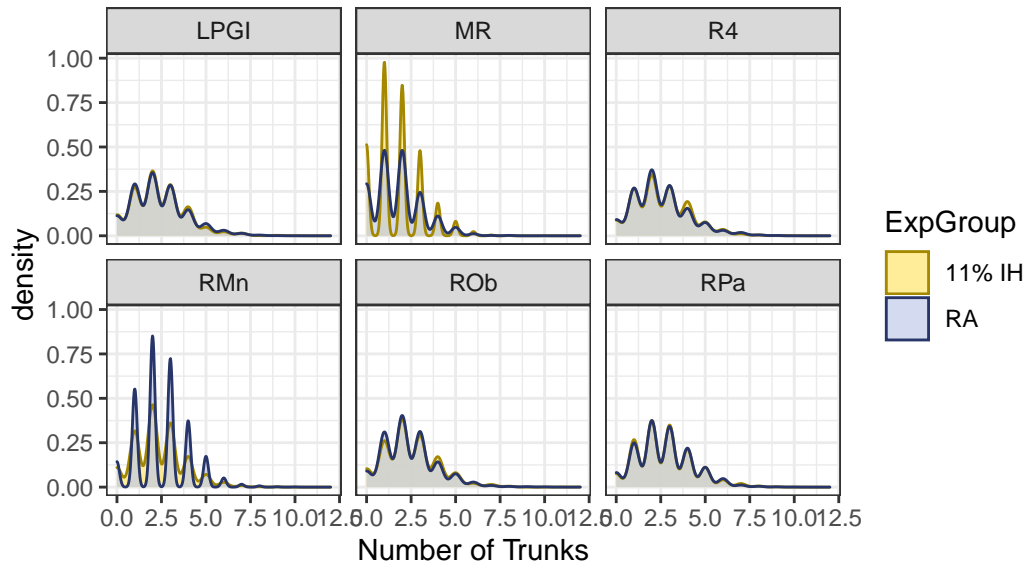
Source: Batch 3

Distribution of Number of Trunk Branches by Region, treatment



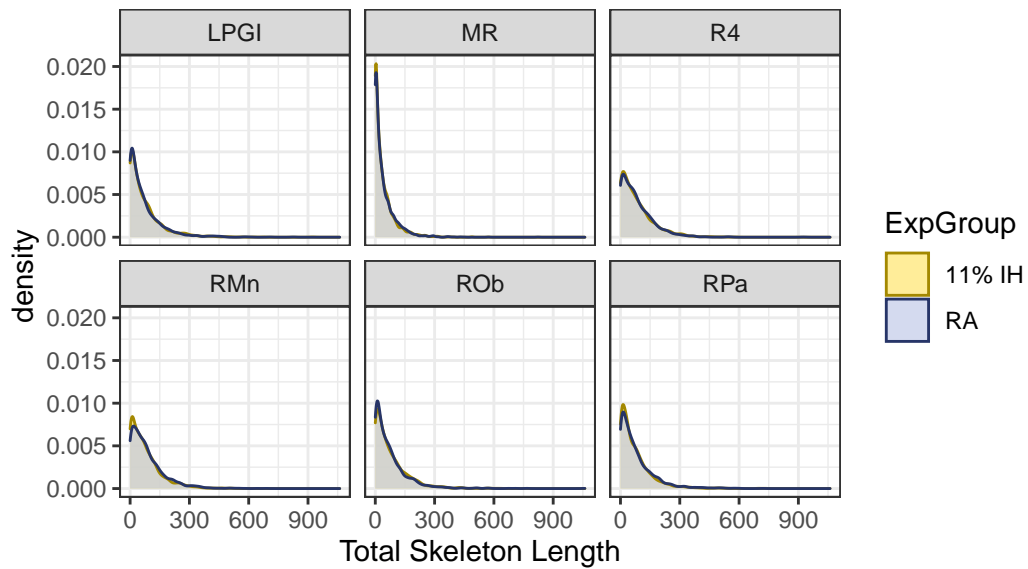
Source: Batch 3

Distribution of Number of Trunks by Region, treatment



Source: Batch 3

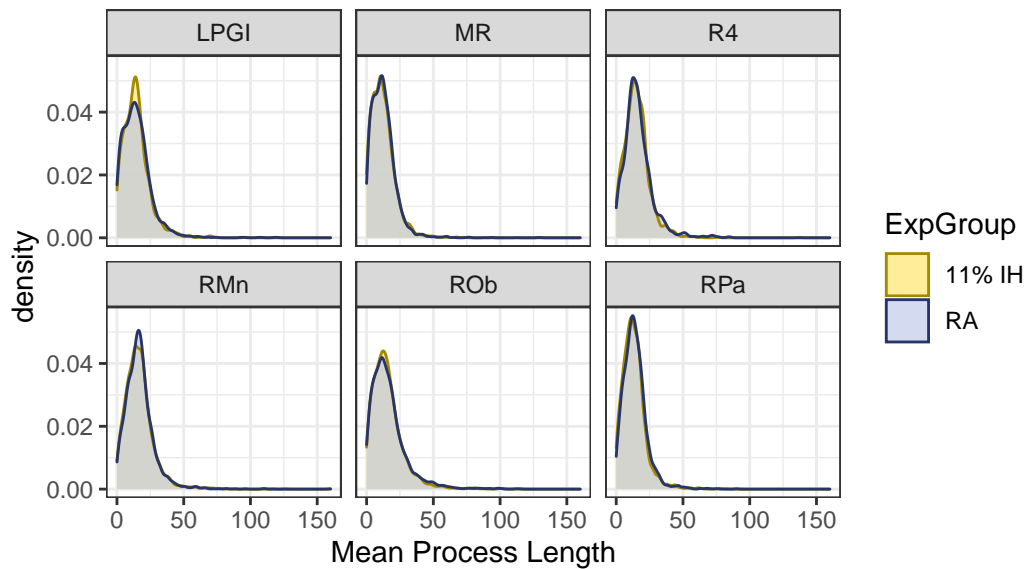
Distribution of Total Skeleton Length by Region, treatment



Source: Batch 3

Warning: Removed 2637 rows containing non-finite values (stat_density).
 Removed 2637 rows containing non-finite values (stat_density).

Distribution of Mean Process Length by Region, treatment



Source: Batch 3

Make summary table (n counts for mouse, region, treatment)

```
tally_df <- sc_df_combined %>%
  group_by(Image_Metadata_Mouse, Image_Metadata_Region,
    ExpGroup, ImageNumber) %>%
  tally(name = "n_cells") %>%
  select(-ImageNumber) %>%
  group_by(Image_Metadata_Mouse, Image_Metadata_Region,
    ExpGroup) %>%
  add_tally(name = "n_images") %>%
  summarize(n_cells = sum(n_cells),
    n_images = median(n_images)) %>%
  mutate(n_cells_per_image = n_cells/n_images)
```

`summarise()` has grouped output by 'Image_Metadata_Mouse',
'Image_Metadata_Region'. You can override using the `groups` argument.

```
kable(tally_df)
```

Image_Metadata_Mouse	Image_Metadata_Regio	ExpGroup	n_cells	n_images	n_cells_per_image
Q1	LPGI	11% IH	80	3	26.66667
Q1	MR	11% IH	468	4	117.00000
Q1	R4	11% IH	91	4	22.75000
Q1	RMn	11% IH	158	3	52.66667
Q1	ROb	11% IH	51	2	25.50000
Q1	RPa	11% IH	190	3	63.33333
Q3	LPGI	11% IH	83	3	27.66667
Q3	MR	11% IH	283	4	70.75000
Q3	R4	11% IH	81	4	20.25000
Q3	RMn	11% IH	186	4	46.50000
Q3	ROb	11% IH	45	1	45.00000
Q3	RPa	11% IH	194	3	64.66667
Q7	LPGI	11% IH	117	4	29.25000
Q7	MR	11% IH	324	4	81.00000
Q7	R4	11% IH	110	4	27.50000
Q7	RMn	11% IH	235	4	58.75000
Q7	ROb	11% IH	91	3	30.33333
Q7	RPa	11% IH	183	4	45.75000
S4	LPGI	RA	125	4	31.25000
S4	MR	RA	214	4	53.50000
S4	R4	RA	90	4	22.50000
S4	RMn	RA	156	4	39.00000
S4	ROb	RA	126	4	31.50000
S4	RPa	RA	229	4	57.25000
S5	LPGI	RA	114	4	28.50000
S5	MR	RA	122	3	40.66667
S5	R4	RA	139	4	34.75000
S5	RMn	RA	252	4	63.00000
S5	ROb	RA	179	5	35.80000
S5	RPa	RA	249	4	62.25000
S6	MR	RA	300	4	75.00000
S6	R4	RA	78	4	19.50000
S6	RMn	RA	265	5	53.00000
S7	LPGI	RA	96	4	24.00000
S7	MR	RA	269	4	67.25000
S7	R4	RA	88	5	17.60000
S7	RMn	RA	147	4	36.75000
S7	ROb	RA	162	4	40.50000
S7	RPa	RA	205	3	68.33333
T5	LPGI	11% IH	120	4	30.00000
T5	MR	11% IH	263	4	65.75000
T5	R4	11% IH	74	4	18.50000
T5	RMn	11% IH	178	4	44.50000
T5	ROb	11% IH	125	4	31.25000
T5	RPa	11% IH	185	4	46.25000
T7	LPGI	11% IH	110	4	27.50000
T7	MR	11% IH	345	4	86.25000

Image_Metadata_Mouse	Image_Metadata_Regio	ExpGroup	n_cells	n_images	n_cells_per_image
T7	R4	11% IH	140	4	35.00000
T7	RMn	11% IH	197	4	49.25000
T7	ROb	11% IH	103	3	34.33333
T7	RPa	11% IH	206	4	51.50000
T8	LPGL	11% IH	97	4	24.25000
T8	MR	11% IH	325	5	65.00000
T8	R4	11% IH	142	5	28.40000
T8	RMn	11% IH	216	4	54.00000
T8	ROb	11% IH	155	4	38.75000
T8	RPa	11% IH	283	5	56.60000
U1	LPGL	11% IH	118	4	29.50000
U1	MR	11% IH	203	4	50.75000
U1	R4	11% IH	139	5	27.80000
U1	RMn	11% IH	167	4	41.75000
U1	ROb	11% IH	158	4	39.50000
U1	RPa	11% IH	193	3	64.33333
U10	LPGL	11% IH	118	4	29.50000
U10	MR	11% IH	281	4	70.25000
U10	R4	11% IH	40	2	20.00000
U10	RMn	11% IH	148	4	37.00000
U10	ROb	11% IH	153	4	38.25000
U10	RPa	11% IH	228	4	57.00000
U4	LPGL	11% IH	126	4	31.50000
U4	MR	11% IH	301	4	75.25000
U4	R4	11% IH	74	5	14.80000
U4	RMn	11% IH	169	4	42.25000
U4	ROb	11% IH	202	4	50.50000
U4	RPa	11% IH	183	3	61.00000
U6	LPGL	11% IH	102	4	25.50000
U6	MR	11% IH	316	4	79.00000
U6	R4	11% IH	159	4	39.75000
U6	RMn	11% IH	266	4	66.50000
U6	ROb	11% IH	149	4	37.25000
U6	RPa	11% IH	209	4	52.25000
V4	LPGL	RA	131	4	32.75000
V4	MR	RA	219	4	54.75000
V4	R4	RA	69	4	17.25000
V4	RMn	RA	194	4	48.50000
V4	ROb	RA	235	4	58.75000
V4	RPa	RA	222	4	55.50000
V5	LPGL	RA	104	4	26.00000
V5	MR	RA	222	4	55.50000
V5	R4	RA	105	4	26.25000
V5	RMn	RA	197	4	49.25000
V5	ROb	RA	155	4	38.75000
V5	RPa	RA	130	3	43.33333
V6	LPGL	RA	114	4	28.50000

Image_Metadata_Mouse	Image_Metadata_Regio	ExpGroup	n_cells	n_images	n_cells_per_image
V6	MR	RA	283	4	70.75000
V6	R4	RA	140	5	28.00000
V6	RMn	RA	199	4	49.75000
V6	ROb	RA	169	4	42.25000
V6	RPa	RA	239	4	59.75000
W2	LPGL	11% IH	126	4	31.50000
W2	MR	11% IH	222	4	55.50000
W2	R4	11% IH	110	4	27.50000
W2	RMn	11% IH	237	4	59.25000
W2	ROb	11% IH	198	4	49.50000
W2	RPa	11% IH	155	3	51.66667
W4	LPGL	11% IH	115	4	28.75000
W4	MR	11% IH	146	4	36.50000
W4	R4	11% IH	152	4	38.00000
W4	RMn	11% IH	210	4	52.50000
W4	ROb	11% IH	136	4	34.00000
W4	RPa	11% IH	222	4	55.50000
W5	LPGL	11% IH	123	4	30.75000
W5	MR	11% IH	305	4	76.25000
W5	R4	11% IH	128	5	25.60000
W5	RMn	11% IH	169	4	42.25000
W5	ROb	11% IH	85	3	28.33333
W5	RPa	11% IH	234	4	58.50000
W9	LPGL	11% IH	103	4	25.75000
W9	MR	11% IH	336	4	84.00000
W9	R4	11% IH	122	4	30.50000
W9	RMn	11% IH	214	4	53.50000
W9	ROb	11% IH	138	4	34.50000
W9	RPa	11% IH	247	4	61.75000
X2	LPGL	RA	111	4	27.75000
X2	MR	RA	179	3	59.66667
X2	R4	RA	123	4	30.75000
X2	RMn	RA	242	4	60.50000
X2	ROb	RA	121	4	30.25000
X2	RPa	RA	183	4	45.75000
X3	LPGL	RA	133	4	33.25000
X3	MR	RA	366	6	61.00000
X3	R4	RA	87	4	21.75000
X3	RMn	RA	226	4	56.50000
X3	ROb	RA	178	4	44.50000
X3	RPa	RA	228	4	57.00000
X5	LPGL	RA	167	5	33.40000
X5	MR	RA	250	5	50.00000
X5	R4	RA	151	4	37.75000
X5	RMn	RA	261	4	65.25000
X5	ROb	RA	117	3	39.00000
X5	RPa	RA	249	5	49.80000

Image_Metadata_Mouse	Image_Metadata_Regio	ExpGroup	n_cells	n_images	n_cells_per_image
Y4	LPGI	RA	135	5	27.00000
Y4	MR	RA	335	4	83.75000
Y4	R4	RA	173	4	43.25000
Y4	RMn	RA	212	4	53.00000
Y4	ROb	RA	137	4	34.25000
Y4	RPa	RA	218	4	54.50000
Y5	LPGI	RA	97	4	24.25000
Y5	MR	RA	213	3	71.00000
Y5	R4	RA	127	4	31.75000
Y5	RMn	RA	186	4	46.50000
Y5	ROb	RA	178	5	35.60000
Y5	RPa	RA	204	4	51.00000
Y6	LPGI	RA	116	4	29.00000
Y6	MR	RA	339	4	84.75000
Y6	R4	RA	154	5	30.80000
Y6	RMn	RA	198	4	49.50000
Y6	ROb	RA	138	4	34.50000
Y6	RPa	RA	177	4	44.25000
Y7	LPGI	RA	160	5	32.00000
Y7	R4	RA	12	1	12.00000
Y7	RMn	RA	135	2	67.50000
Y7	ROb	RA	196	5	39.20000
Y7	RPa	RA	187	4	46.75000

```

# total_count_RA <- sc_df_combined %>%
# filter(ExpGroup=='RA') %>% count() %>% pull()
# total_count_IH <- sc_df_combined %>%
# filter(ExpGroup=='11% IH') %>% count() %>% pull()

```