Cre-IHC-colocalization

CT Berezin

2022-11-02

```
setwd("D:/Microscopy/IHC images/20221020 - Opn4-cre-EGFP/analysis/processing/CSVs")

#wm 5 and 7 could have more GFP cells
wdir <- getwd()

count_list <- list.files("D:/Microscopy/IHC images/20221020 - Opn4-cre-EGFP/analysis/processing/CSVs", p
length(count_list)
```

```
## [1] 10
```

```
coloc_list <- list.files("D:/Microscopy/IHC images/20221020 - Opn4-cre-EGFP/analysis/processing/CSVs", p
length(coloc_list)
```

```
## [1] 10
```

```
img_sumstats <- data.frame()
mouse_sumstats <- data.frame()
dapi_names <- c()
gfp_names <- c()
cre_names <- c()

for (file in count_list){
  counts <- read.csv(paste0(wdir,"/",file))
  counts <- counts %>%
    rename(VolUnit = Vol..unit.,
           VolPix = Vol..pix.) %>%
    select(c(Nb, Label, Name, VolUnit, VolPix))
  counts <- counts %>% mutate(Name = str_replace(Name, "obj325-val3", "obj325-val1"),
                              Name = str_replace(Name, "obj532-val2", "obj532-val1"),
                              Name = str_replace(Name, "obj802-val2", "obj801-val1"))
  cutoffs <- counts %>% filter(str_detect(Name, ".*(val1)$"))
  counts <- counts %>% mutate(type = case_when(
    between(Nb+1, cutoffs[1,1], cutoffs[2,1]) ~ "DAPI",
    between(Nb+1, cutoffs[2,1], cutoffs[3,1]) ~ "GFP",
    Nb+1 > cutoffs[3,1] ~ "cre"
  ))
  dapi_names[file] <- counts %>% filter(type == "DAPI") %>% select(Name)
  gfp_names[file] <- counts %>% filter(type == "GFP") %>% select(Name)
  cre_names[file] <- counts %>% filter(type == "cre") %>% select(Name)
  #print(head(counts))
```

```
  imgsumstats <- counts %>%
    group_by(type) %>%
    summarise(filename = file,
              n = n(),
              minVol = min(VolUnit),
              medVol = median(VolUnit),
              meanVol = mean(VolUnit),
              maxVol = max(VolUnit))
  img_sumstats <- rbind(img_sumstats, imgsumstats)
  mousesumstats <- counts %>% filter(type == "GFP") %>%
    mutate(mouse = as.factor(str_extract(file, "(?<=_).*(?=-[:digit:]_)")))
  mouse_sumstats <- rbind(mouse_sumstats, mousesumstats)
}
img_sumstats
```

```
## # A tibble: 30 x 7
##    type  filename                                n   minVol  medVol meanVol maxVol
##    <chr> <chr>                               <int>    <dbl>   <dbl>   <dbl>  <dbl>
##  1 cre   M_M-6-40x-8avg-1_quant.csv         4744   0.0243   0.292   1.07   505.
##  2 DAPI  M_M-6-40x-8avg-1_quant.csv          375   0.243    1.78    3.41    52.2
##  3 GFP   M_M-6-40x-8avg-1_quant.csv            2 261.     290.    290.     319.
##  4 cre   M_M-6-40x-8avg-2_quant-fixed.csv   664   0.0243   0.243   1.03   205.
##  5 DAPI  M_M-6-40x-8avg-2_quant-fixed.csv   322   0.243    1.96    3.98   154.
##  6 GFP   M_M-6-40x-8avg-2_quant-fixed.csv     4  34.1    244.    224.     373.
##  7 cre   M_M-6-40x-8avg-3_quant-fixed.csv  1066   0.0243   0.219   0.788  199.
##  8 DAPI  M_M-6-40x-8avg-3_quant-fixed.csv  1063   0.243    1.27    3.24   101.
##  9 GFP   M_M-6-40x-8avg-3_quant-fixed.csv     3 139.     155.    164.     198.
## 10 cre   M_M-6-40x-8avg-4_quant-fixed.csv  1933   0.0243   0.268   0.819  121.
## # ... with 20 more rows
```

```
mouse_gfp_sumstats <- mouse_sumstats %>% group_by(mouse) %>%
    summarise(minVol = min(VolUnit),
              medVol = median(VolUnit),
              meanVol = mean(VolUnit),
              maxVol = max(VolUnit))
mouse_gfp_sumstats
```

```
## # A tibble: 2 x 5
##   mouse        minVol medVol meanVol maxVol
##   <fct>         <dbl>  <dbl>   <dbl>  <dbl>
## 1 M-6-40x-8avg   34.1   211.    204.   373.
## 2 wm-40x-8avg    48.4   210.    222.   424.
```

```
overall_sumstats <- mouse_sumstats %>%
    summarise(minVol = min(VolUnit),
              medVol = median(VolUnit),
              meanVol = mean(VolUnit),
              maxVol = max(VolUnit))
overall_sumstats
```

```
##     minVol   medVol  meanVol   maxVol
## 1 34.11097 210.2982 212.8922 424.2459
```

```
names(dapi_names) <- sub("_quant.*", "", names(dapi_names))
names(dapi_names) <- sub("M_", "", names(dapi_names))
names(gfp_names) <- sub("_quant.*", "", names(gfp_names))
names(gfp_names) <- sub("M_", "", names(gfp_names))
names(cre_names) <- sub("_quant.*", "", names(cre_names))
names(cre_names) <- sub("M_", "", names(cre_names))

#gfp_names
```

```
coloc_sumstats <- data.frame()

for (file in coloc_list){
  coloc_count <- read.csv(paste0(wdir,"/",file))
  filename <- paste0(file)
  filename <- str_replace(filename, "C_", "")
  filename <- str_replace(filename, "_cre.*", "")
  coloc_count <- coloc_count %>% mutate(Label1 = str_replace(Label1, "obj325-val3", "obj325-val1"),
                            Label1 = str_replace(Label1, "obj532-val2", "obj532-val1"),
                            Label1 = str_replace(Label1, "obj802-val2", "obj801-val1"))
  coloc_count <- coloc_count %>%
    select(c(Label1, Label2, coloc, PcColoc)) %>%
    filter(Label1 %in% gfp_names[filename][[1]] & Label2 %in% cre_names[filename][[1]])
  #coloc_count <- full_join(coloc_count, cell_types, by=c("Label1"="Name"))
  #coloc_count <- full_join(coloc_count, cell_types, by=c("Label2"="Name"), suffix=c("_1","_2"))
  #coloc_count <- coloc_count %>% filter(type_1 == "GFP" & type_2 == "cre", na.rm = TRUE)
  sumstats <- coloc_count %>% group_by(Label1) %>%
      summarise(filename = file,
              n = n(),
              colocPerc = sum(PcColoc),
              trueColoc1 = colocPerc >= 1,
              trueColoc5 = colocPerc >= 5,
              posObjs = sum(coloc > 1),
              truePos1 = posObjs >= 1,
              truePos5 = posObjs >= 5,
              )
  coloc_sumstats <- rbind(coloc_sumstats, sumstats)
}

head(coloc_sumstats)
```

```
## # A tibble: 6 x 9
##   Label1      filename         n coloc~1 trueC~2 trueC~3 posObjs trueP~4 trueP~5
##   <chr>       <chr>        <int>   <dbl> <lgl>   <lgl>     <int> <lgl>   <lgl>
## 1 obj376-val1 C_M-6-40x-8~  4744   6.17  TRUE    TRUE         23 TRUE    TRUE
## 2 obj377-val2 C_M-6-40x-8~  4744   3.58  TRUE    FALSE        12 TRUE    TRUE
## 3 obj325-val1 C_M-6-40x-8~   663   0.637 FALSE   FALSE         3 TRUE    FALSE
## 4 obj326-val4 C_M-6-40x-8~   663  78.7   TRUE    TRUE          1 TRUE    FALSE
## 5 obj327-val5 C_M-6-40x-8~   663   1.08  TRUE    FALSE        11 TRUE    TRUE
## 6 obj329-val7 C_M-6-40x-8~   663  12.3   TRUE    TRUE          4 TRUE    FALSE
## # ... with abbreviated variable names 1: colocPerc, 2: trueColoc1,
## #   3: trueColoc5, 4: truePos1, 5: truePos5
```

```
coloc_stats <- coloc_sumstats %>% group_by(filename) %>% summarise(n = n(),
                                                                   ">1% coloc" = sum(trueColoc1)/n*100,
                                                                   ">5% coloc" = sum(trueColoc5)/n*100,
                                                                   minPerc = min(colocPerc),
                                                                   meanPerc = mean(colocPerc),
                                                                   maxPerc = max(colocPerc),
                                                                   ">1dot" = sum(truePos1)/n*100,
                                                                   ">5dots" = sum(truePos5)/n*100,)
kable(coloc_stats)
```

filename	n	>1% coloc	>5% coloc	minPerc	meanPerc	maxPerc	>1dot	>5dots
C_M-6-40x-8avg-1_cre-gfp-colocalization.csv	2	100.00000	50.00000	3.5785137	4.8746999	6.1708861	100	100.00000
C_M-6-40x-8avg-2_cre-gfp-colocalization-fixed.csv	4	75.00000	50.00000	0.6374881	23.1639090	78.6729363	100	25.00000
C_M-6-40x-8avg-3_cre-gfp-colocalization-fixed.csv	3	66.66667	66.66667	0.1099592	33.9072441	87.3190302	100	33.33333
C_M-6-40x-8avg-4_cre-gfp-colocalization-fixed.csv	3	100.00000	100.00000	12.3486683	31.4440396	52.9235382	100	33.33333
C_M-6-40x-8avg-5_cre-gfp-colocalization-fixed.csv	1	0.00000	0.00000	0.0000000	0.0000000	0.0000000	0	0.00000
C_wm-40x-8avg-1_cre-gfp-colocalization.csv	5	100.00000	100.00000	9.8334474	20.0049178	32.8857101	100	100.00000
C_wm-40x-8avg-5_cre-gfp-colocalization-fixed.csv	2	100.00000	50.00000	3.0524605	9.7003751	16.3482896	100	50.00000
C_wm-40x-8avg-6_cre-gfp-colocalization.csv	3	100.00000	66.66667	3.0672556	5.6041485	7.6546925	100	100.00000
C_wm-40x-8avg-7_cre-gfp-colocalization.csv	2	100.00000	100.00000	31.3742819	45.2600053	59.1457286	100	50.00000
C_wm-40x-8avg-8_cre-gfp-colocalization.csv	1	0.00000	0.00000	0.0770416	0.0770416	0.0770416	100	0.00000

```
coloc5_stats <- coloc_sumstats %>% group_by(filename) %>%
  filter(colocPerc > 5) %>%
  summarise(n = n(),
            ">5% coloc" = sum(trueColoc5)/n*100,
            minPerc = min(colocPerc),
            meanPerc = mean(colocPerc),
            maxPerc = max(colocPerc))
kable(coloc5_stats)
```

filename	n	>5% coloc	minPerc	meanPerc	maxPerc
C_M-6-40x-8avg-1_cre-gfp-colocalization.csv	1	100	6.170886	6.170886	6.170886
C_M-6-40x-8avg-2_cre-gfp-colocalization-fixed.csv	2	100	12.268188	45.470562	78.672936
C_M-6-40x-8avg-3_cre-gfp-colocalization-fixed.csv	2	100	14.292743	50.805887	87.319030

filename	n	>5% coloc	minPerc	meanPerc	maxPerc
C_M-6-40x-8avg-4_cre-gfp-colocalization-fixed.csv	3	100	12.348668	31.444040	52.923538
C_wm-40x-8avg-1_cre-gfp-colocalization.csv	5	100	9.833447	20.004918	32.885710
C_wm-40x-8avg-5_cre-gfp-colocalization-fixed.csv	1	100	16.348290	16.348290	16.348290
C_wm-40x-8avg-6_cre-gfp-colocalization.csv	2	100	6.090497	6.872595	7.654693
C_wm-40x-8avg-7_cre-gfp-colocalization.csv	2	100	31.374282	45.260005	59.145729

```
mouse_coloc_sumstats <- coloc_sumstats %>%
  mutate(mouse = str_extract(filename, "(?<=C_).*(?=-[:digit:]_)")) %>%
  filter(colocPerc > 5) %>%
  group_by(mouse) %>%
  summarise(
     n = n(),
     minPerc = min(colocPerc),
     meanPerc = mean(colocPerc),
     maxPerc = max(colocPerc)
  )
mouse_coloc_sumstats
```

```
## # A tibble: 2 x 5
##   mouse            n minPerc meanPerc maxPerc
##   <chr>        <int>   <dbl>    <dbl>   <dbl>
## 1 M-6-40x-8avg     8    6.17     36.6    87.3
## 2 wm-40x-8avg     10    6.09     22.1    59.1
```

```
overall_coloc_sumstats <- coloc_sumstats %>%
  filter(colocPerc > 5) %>%
  summarise(
     n = n(),
     minPerc = min(colocPerc),
     meanPerc = mean(colocPerc),
     maxPerc = max(colocPerc)
  )
overall_coloc_sumstats
```

```
## # A tibble: 1 x 4
##       n minPerc meanPerc maxPerc
##   <int>   <dbl>    <dbl>   <dbl>
## 1    18    6.09     28.5    87.3
```