Cre-IHC-colocalization

CT Berezin

2022-11-02

```
setwd("D:/Microscopy/IHC images/20221019 - cre validation brain/processing/CSVs")
wdir <- getwd()

dapi_list <- list.files(wdir, pattern=".*dapi-quant.*")
length(dapi_list)
```

```
## [1] 12
```

```
cre_list <- list.files(wdir, pattern=".*cre-quant.*")
length(cre_list)
```

```
## [1] 12
```

```
coloc_list <- list.files(wdir, pattern=".*coloc.*")
length(coloc_list)
```

```
## [1] 24
```

```
dapi_sumstats <- data.frame()
dapi__sumstats <- data.frame()
dapi_names <- c()

for (file in dapi_list){
  dapi_count <- read.csv(paste0(wdir,"/",file))
  dapi_count <- dapi_count %>%
    rename(VolUnit = Vol..unit.,
           VolPix = Vol..pix.) %>%
    select(c(Nb, Name, VolUnit, VolPix))
  dapi_names[file] <- dapi_count %>% select(Name)
  #print(head(dapi_count))
  sumstats <- dapi_count %>%
    summarise(filename = file,
              n = n(),
              minVol = min(VolUnit),
              medVol = median(VolUnit),
              meanVol = mean(VolUnit),
              maxVol = max(VolUnit))
  dapi_sumstats <- rbind(dapi_sumstats, sumstats)
  mousesumstats <- dapi_count %>%
```

```
    mutate(region = as.factor(str_extract(file, "(?<=McKO-).*(?=-section)")),
           mouse = "McKO")
  dapi__sumstats <- rbind(dapi__sumstats, mousesumstats)
}

dapi_region_sumstats <- dapi__sumstats %>% group_by(region) %>%
    summarise(minVol = min(VolUnit),
              medVol = median(VolUnit),
              meanVol = mean(VolUnit),
              maxVol = max(VolUnit))
dapi_region_sumstats
```

```
## # A tibble: 3 x 5
##   region minVol medVol meanVol maxVol
##   <fct>   <dbl>  <dbl>   <dbl>  <dbl>
## 1 CP      0.535   378.    408.  1480.
## 2 HY      0.973   319.    326.  1034.
## 3 TH      1.07    360.    374.  1226.
```

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

```
## # A tibble: 1 x 5
##   mouse minVol medVol meanVol maxVol
##   <chr>  <dbl>  <dbl>   <dbl>  <dbl>
## 1 McKO   0.535   341.    361.  1480.
```

```
names(dapi_names) <- sub("_dapi.*", "", names(dapi_names))
names(dapi_names) <- sub("M_", "", names(dapi_names))

cre_sumstats <- data.frame()
cre_names <- c()

for (file in cre_list){
  cre_count <- read.csv(paste0(wdir,"/",file))
  cre_count <- cre_count %>%
    rename(VolUnit = Vol..unit.,
           VolPix = Vol..pix.) %>%
    select(c(Nb, Name, VolUnit, VolPix))
  cre_names[file] <- cre_count %>% select(Name)
  #print(head(cre_count))
  sumstats <- cre_count %>%
    summarise(filename = file,
              n = n(),
              minVol = min(VolUnit),
              medVol = median(VolUnit),
              meanVol = mean(VolUnit),
              maxVol = max(VolUnit))
```

```
  cre_sumstats <- rbind(cre_sumstats, sumstats)
}
head(cre_sumstats)
```

```
##                                          filename    n     minVol    medVol
## 1      M_McKO-CP-section1-L-40x-4avg-1_cre-quant.csv 1170 0.02433022 0.6082555
## 2      M_McKO-CP-section1-R-40x-4avg-1_cre-quant.csv 1255 0.02433022 0.6082555
## 3      M_McKO-CP-section9-L-40x-4avg-1_cre-quant.csv  616 0.02433022 0.7055763
## 4      M_McKO-CP-section9-R-40x-4avg-1_cre-quant.csv  685 0.02433022 0.6325857
## 5 M_McKO-HY-section1-VLPO-L-40x-4avg-1_cre-quant.csv 1276 0.02433022 0.8515576
## 6 M_McKO-HY-section2-VLPO-L-40x-4avg-1_cre-quant.csv 1996 0.02433022 0.5595950
##    meanVol   maxVol
## 1 1.387426 47.34661
## 2 1.720621 68.87885
## 3 2.278628 40.55847
## 4 1.910934 27.59047
## 5 2.405298 44.91358
## 6 1.599980 45.08389
```

```
names(cre_names) <- sub("_cre.*", "", names(cre_names))
names(cre_names) <- sub("M_", "", names(cre_names))
```

```
if(file.exists(wdir, pattern=".*filtered.csv")){
  NULL
} else {
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.*", "")
  #print(dapi_names[filename])
  #print(dapi_names[filename][[1]])
  coloc_count <- coloc_count %>%
    select(c(Nb, Obj1, Obj2, Label1, Label2, coloc, PcColoc)) %>%
    filter(Label1 %in% dapi_names[filename][[1]] & Label2 %in% cre_names[filename][[1]])
  write.csv(coloc_count, file=paste0(wdir,"/",file,"-filtered.csv"))
  }
}
```

```
## Warning in if (file.exists(wdir, pattern = ".*filtered.csv")) {: the condition
## has length > 1 and only the first element will be used
```

```
## NULL
```

```
coloc_list_filt <- list.files(wdir, pattern=".*filtered.csv")
length(coloc_list_filt)
```

```
## [1] 12
```

```
coloc_sumstats <- data.frame()

for (file in coloc_list_filt){
    coloc_count <- read.csv(paste0(wdir,"/",file))
    sumstats <- coloc_count %>% group_by(Label1) %>%
      summarise(filename = file,
                n = n(),
                colocPerc = sum(PcColoc),
                trueColoc1 = colocPerc > 1,
                trueColoc5 = colocPerc > 5,
                pos = sum(coloc > 1),
                truePos1 = pos > 1,
                truePos5 = pos > 5,
                min = min(coloc),
                median = median(coloc),
                mean = mean(coloc),
                max = max(coloc))
  coloc_sumstats <- rbind(coloc_sumstats, sumstats)
}

coloc_stats <- coloc_sumstats %>% group_by(filename) %>% summarise(n = n(),
                                                                   pos1Perc = sum(truePos1)/n*100,
                                                                   pos5Perc = sum(truePos5)/n*100,
                                                                   perc1Coloc = sum(trueColoc1)/n*100,
                                                                   perc5Coloc = sum(trueColoc5)/n*100)

kable(coloc_stats)
```

filename	n	pos1Perc	pos5Perc	perc1Coloc	perc5Coloc
C_McKO-CP-section1-L-40x-4avg-1_cre-colocalization.csv-filtered.csv	267	34.08240	1.8726592	8.239700	2.2471910
C_McKO-CP-section1-R-40x-4avg-1_cre-colocalization.csv-filtered.csv	210	56.19048	11.904761	22.380952	3.3333333
C_McKO-CP-section9-L-40x-4avg-1_cre-colocalization.csv-filtered.csv	173	39.30636	2.3121387	13.294798	0.5780347
C_McKO-CP-section9-R-40x-4avg-1_cre-colocalization.csv-filtered.csv	198	41.91919	3.5353535	9.595960	0.5050505
C_McKO-HY-section1-VLPO-L-40x-4avg-1_cre-colocalization.csv-filtered.csv	191	41.88482	9.4240838	19.895288	1.0471204
C_McKO-HY-section2-VLPO-L-40x-4avg-1_cre-colocalization.csv-filtered.csv	263	51.33080	15.209125	20.532319	2.2813688
C_McKO-HY-section2-VLPO-R-40x-4avg-1_cre-colocalization.csv-filtered.csv	240	64.58333	12.500000	24.583333	0.8333333
C_McKO-HY-section5-SCN-L-40x-4avg-1_cre-colocalization.csv-filtered.csv	208	46.63462	8.1730769	14.423077	1.4423077
C_McKO-HY-section5-SCN-R-40x-4avg-1_cre-colocalization.csv-filtered.csv	237	65.82278	23.206751	31.223629	2.5316456
C_McKO-HY-section6-VLPO-40x-4avg-1_cre-colocalization.csv-filtered.csv	175	26.85714	0.5714286	9.142857	1.1428571
C_McKO-TH-section2-PVT-40x-4avg-1_cre-colocalization.csv-filtered.csv	189	63.49206	26.455026	35.978836	5.8201058
C_McKO-TH-section4-PVT-40x-4avg-1_cre-colocalization.csv-filtered.csv	209	54.54545	21.531100	19.617225	1.9138756

```
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_McKO-CP-section1-L-40x-4avg-1_cre-colocalization.csv-filtered.csv	6	100	5.136068	7.076593	9.433962
C_McKO-CP-section1-R-40x-4avg-1_cre-colocalization.csv-filtered.csv	7	100	5.067116	21.378387	100.000000
C_McKO-CP-section9-L-40x-4avg-1_cre-colocalization.csv-filtered.csv	1	100	5.789817	5.789817	5.789817
C_McKO-CP-section9-R-40x-4avg-1_cre-colocalization.csv-filtered.csv	1	100	5.932444	5.932444	5.932444
C_McKO-HY-section1-VLPO-L-40x-4avg-1_cre-colocalization.csv-filtered.csv	2	100	7.381281	10.094696	12.808110
C_McKO-HY-section2-VLPO-L-40x-4avg-1_cre-colocalization.csv-filtered.csv	6	100	5.089018	6.473186	10.121892
C_McKO-HY-section2-VLPO-R-40x-4avg-1_cre-colocalization.csv-filtered.csv	2	100	5.778710	7.327523	8.876337
C_McKO-HY-section5-SCN-L-40x-4avg-1_cre-colocalization.csv-filtered.csv	3	100	5.170185	5.928846	6.334301
C_McKO-HY-section5-SCN-R-40x-4avg-1_cre-colocalization.csv-filtered.csv	6	100	5.198559	13.353711	47.500000
C_McKO-HY-section6-VLPO-40x-4avg-1_cre-colocalization.csv-filtered.csv	2	100	7.434402	9.475336	11.516270
C_McKO-TH-section2-PVT-40x-4avg-1_cre-colocalization.csv-filtered.csv	11	100	5.280623	15.447178	100.000000
C_McKO-TH-section4-PVT-40x-4avg-1_cre-colocalization.csv-filtered.csv	4	100	5.664105	8.207033	15.225933

```
region_coloc_sumstats <- coloc_sumstats %>%
  mutate(region = as.factor(str_extract(filename, "(?<=McKO-).*(?=-section)"))) %>%
  filter(colocPerc > 5) %>%
  group_by(region) %>%
  summarise(
     n = n(),
     minPerc = min(colocPerc),
     meanPerc = mean(colocPerc),
     maxPerc = max(colocPerc)
  )
region_coloc_sumstats
```

```
## # A tibble: 3 x 5
##   region     n minPerc meanPerc maxPerc
##   <fct>  <int>   <dbl>    <dbl>   <dbl>
```

```
## 1 CP        15    5.07    13.6    100
## 2 HY        21    5.09     9.07    47.5
## 3 TH        15    5.28    13.5    100
```

```
mouse_coloc_sumstats <- coloc_sumstats %>%
  mutate(mouse = "McKO") %>%
  filter(colocPerc > 5) %>%
  group_by(mouse) %>%
  summarise(
     n = n(),
     minPerc = min(colocPerc),
     meanPerc = mean(colocPerc),
     maxPerc = max(colocPerc)
  )
mouse_coloc_sumstats
```

```
## # A tibble: 1 x 5
##   mouse     n minPerc meanPerc maxPerc
##   <chr> <int>   <dbl>    <dbl>   <dbl>
## 1 McKO     51    5.07     11.7     100
```